ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Sprachverarbeitung I / 3 HS 2016

Kurzzeitanalyse: DFT, Autokorrelation

Buch: Kapitel 4.1 – 4.4

Beat Pfister

Sprachverarbeitung I / 3

Vorlesung: Kurzzeitanalyse von Sprachsignalen

- Zweck der Kurzzeitanalyse
- Diskrete Fouriertransformation
- Spektrum periodischer Signale
- Spektrum rauschartiger Signale
- Autokorrelation

Übung: diskrete Fouriertransformation

Kurzzeitanalyse

Beispiel: Spektral- oder Fourieranalyse

	Langzeitanalyse (stationäres Signal)	Kurzzeitanalyse (quasi-stationäres Signal)
Analyseabschnitt:	möglichst lang	kurz, nahezu stationär >>>
erwartetes Resultat:	spektrale Zusammen-setzung des Signals (ein Vektor)	spektrale Zusammensetzung zu bestimmten Zeitpunkten (Sequenz von Vektoren)
Beispiele:	stationäres Signal >>> Sprachsignal >>>	stationäres Signal >>> Sprachsignal >>> >>>

Diskrete Fouriertransformation (DFT)

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j(2\pi/N)kn} \qquad 0 \le k \le N-1$$

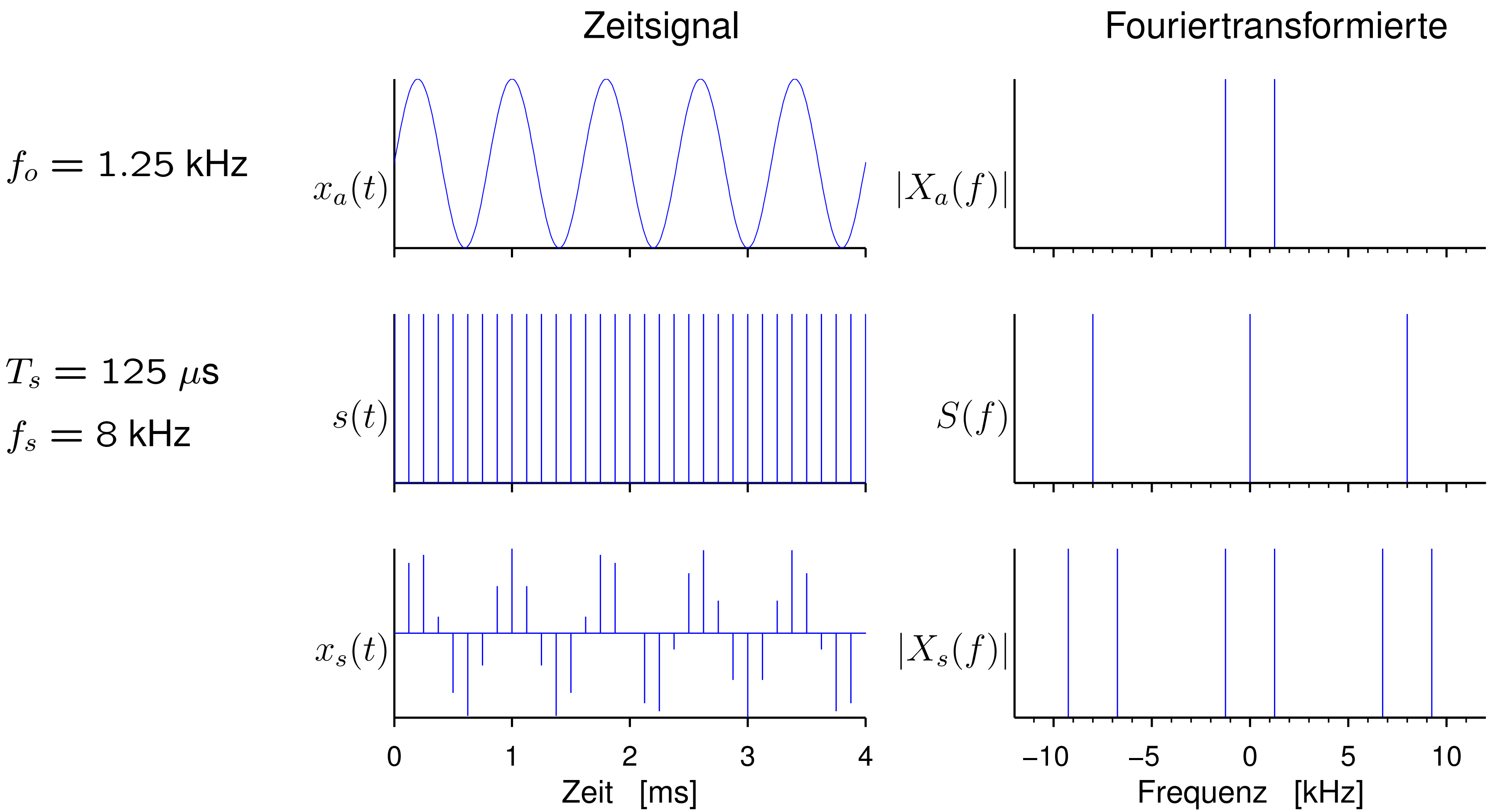
$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j(2\pi/N)kn} \qquad 0 \le n \le N-1$$

Zeitauflösung: $T_s = 1/f_s$

Frequenzauflösung: f_s/N

Zusammenhang zwischen DFT und kontinuierlicher FT ?

Signalabtastung

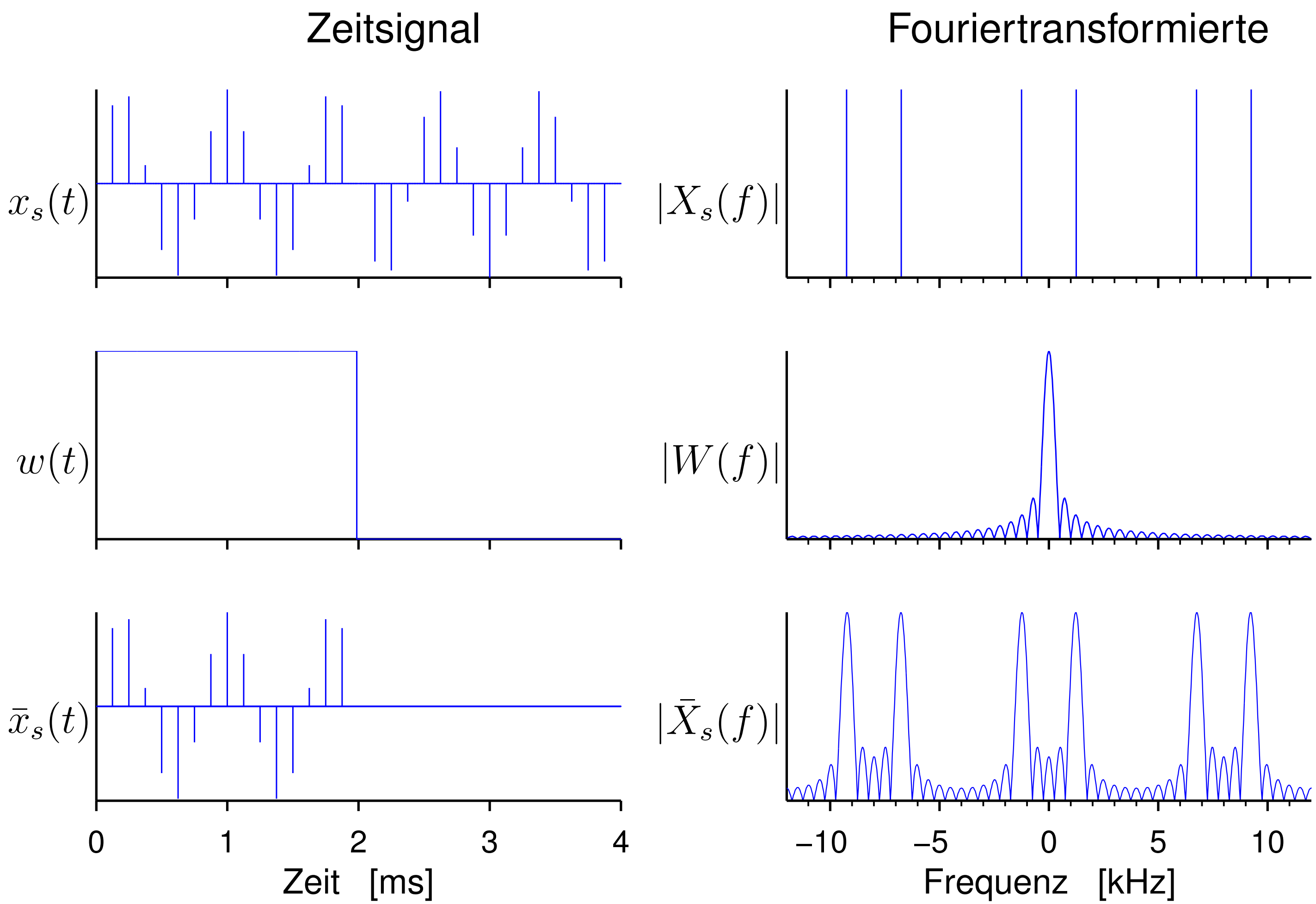


Kontinuierliche vs. diskrete FT — Analysefenster

$f_o = 1.25$ kHz

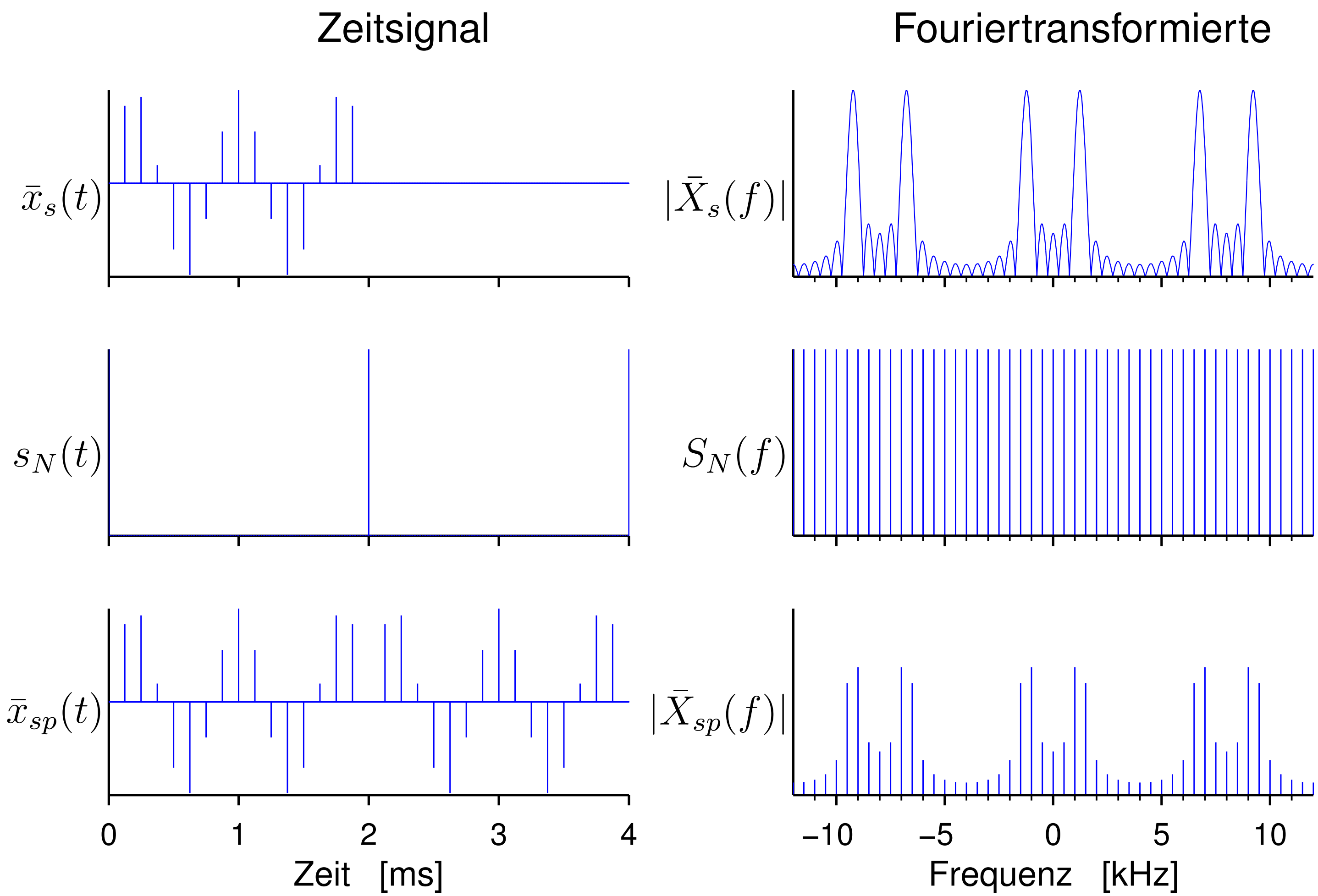
Fenster: 2 ms

($N = 16$)



Kontinuierliche vs. diskrete FT   Spektrumabtastung

$f_o = 1.25$ kHz



$f_s/N = 500$ Hz

Diskrete Fouriertransformation

Achtung! Bei der Anwendung einer N-Punkt-DFT nimmt man implizit an, dass das Zeitsignal $x(n)$ periodisch mit der Länge N ist!

Folge: DFT liefert nur dann die tatsächliche spektrale Zusammensetzung eines Signals, wenn der Analyseabschnitt eine ganze Anzahl Perioden des Signals enthält.

Fragen:

- Wie gross ist der Fehler, falls diese Bedingung nicht zutrifft?
- Was sagt die Fouriertransformierte über das Spektrum aus?

Einfluss der Fensterfunktion auf die DFT

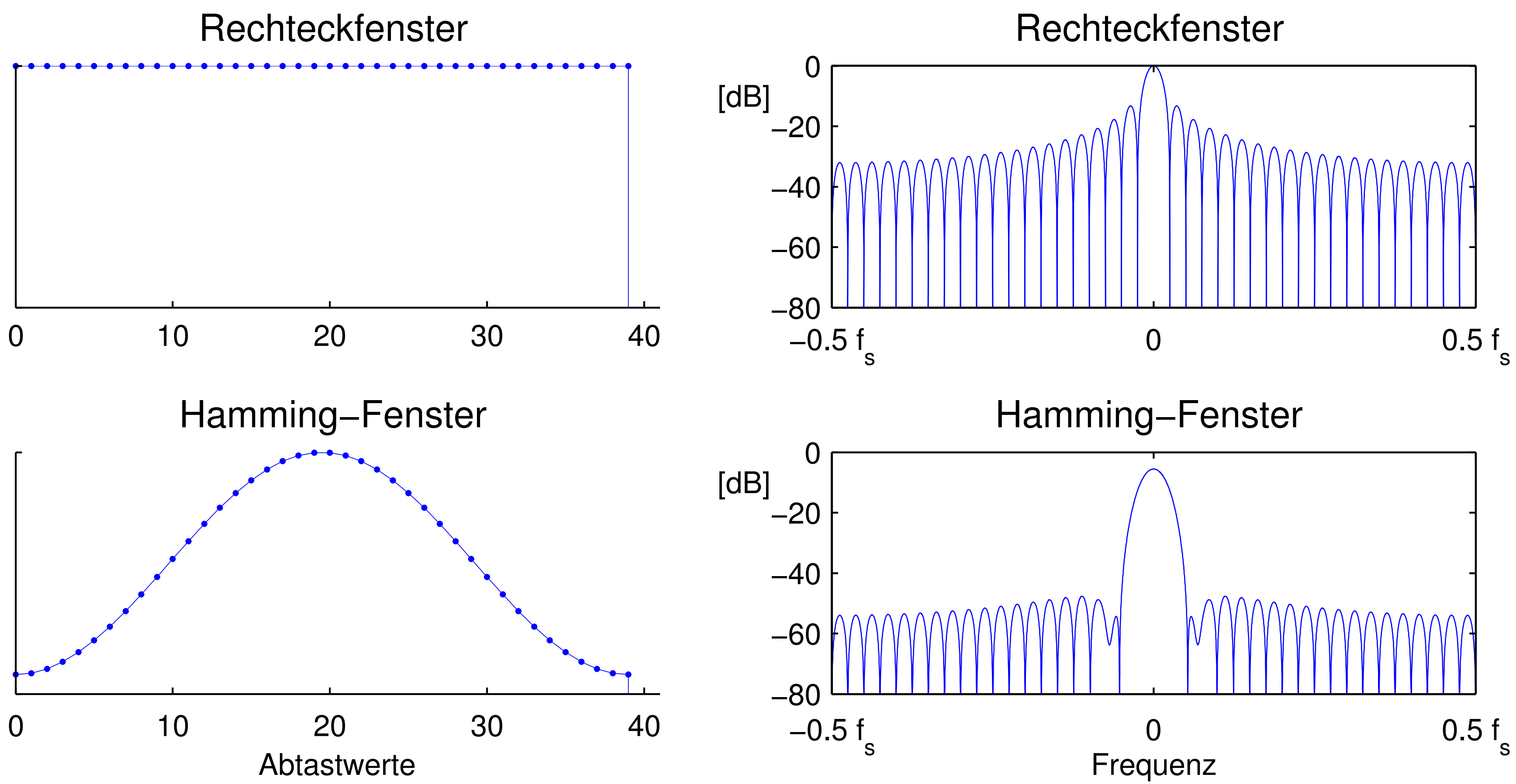
Form und Länge des Analysefensters beeinflussen das Resultat der DFT und damit den Analysefehler.

Multiplikation des Signals mit Fenster $w(n)$ ergibt Faltung des Signalspektrums mit dem Spektrum $W(k)$ des Fensters.

Zwei unerwünschte Effekte:

- Verschmieren: abhängig von Breite des Hauptlappens
- Lecken: abhängig von Höhe der Nebenlappen

Resultat der DFT ist abhängig von Fensterfunktion



>>>

>>>

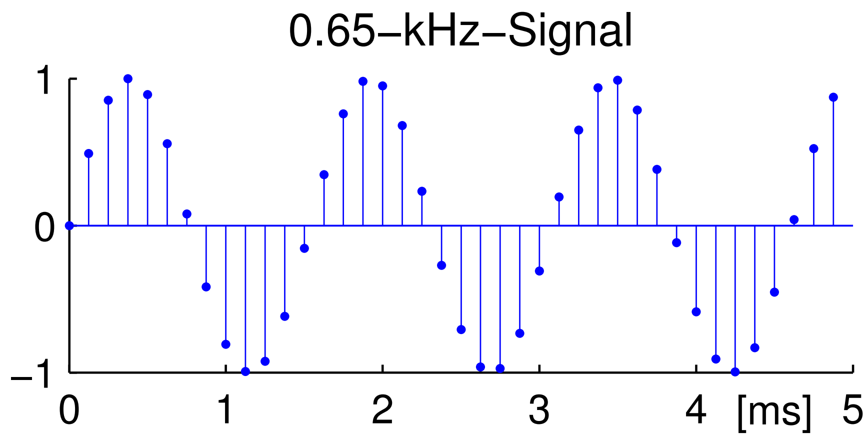
Frequenzauflösung der Fouriertransformation

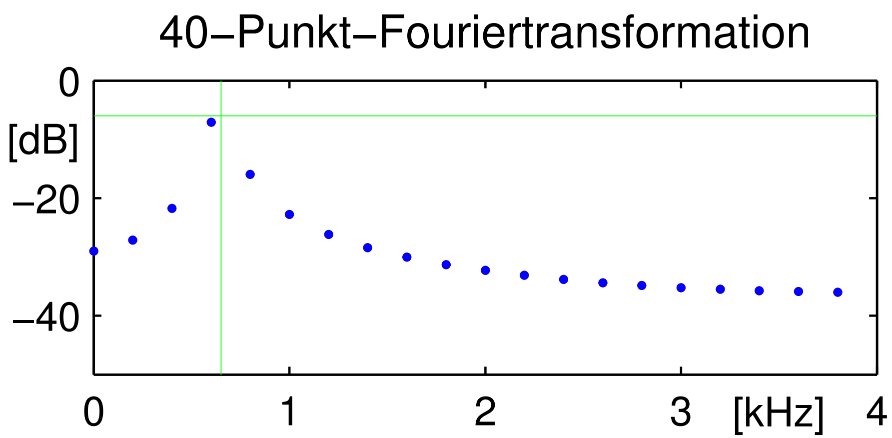
Aufgabe:

Bestimmen der Frequenz
eines 650 Hz Sinus
aus $N = 40$ Abtastwerten
mit Abtastfrequenz $f_s = 8$ kHz

Problem:

Frequenzauflösung $f_s/N = 200$ Hz !





Verbesserung der spektralen Auflösung

Problem: Kurzer Signalabschnitt: $x(n)$, $n = 0, \ldots, N-1$

Folge: Spektrale Auflösung gering, nämlich: f_s/N

(z.B. Sprachsignal mit $f_s = 8$ kHz und $N = 240 \quad \longrightarrow \quad f_\triangle = 33.3$Hz)

Lösung: Verlängerung von $x(n)$ mit N Nullen (*zero padding*):

ergibt die $2N$ lange Sequenz: $x'(n) = x_0, x_1, \ldots, x_{N-1}, 0, 0, \ldots, 0$

$\longrightarrow$ spektrale Auflösung verdoppelt

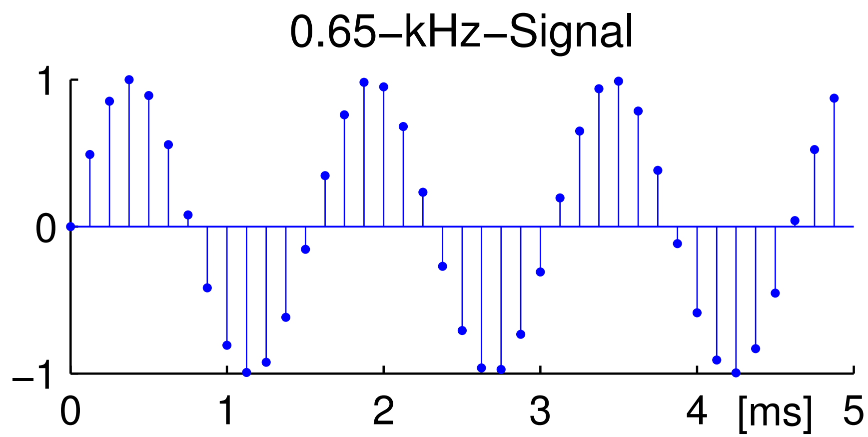
Frage: Zusammenhang zwischen $X(k)$ und $X'(k)$?

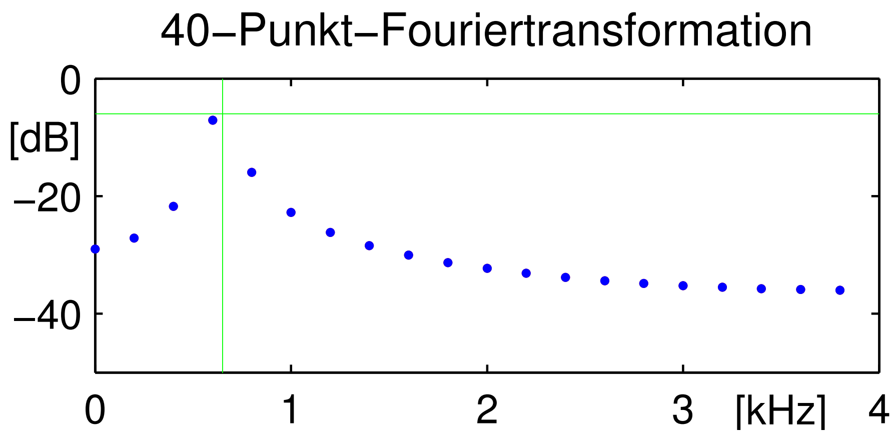
Antwort: $X(k) = X'(2k)$, d.h. ursprüngliche Punkte unverändert, weil:

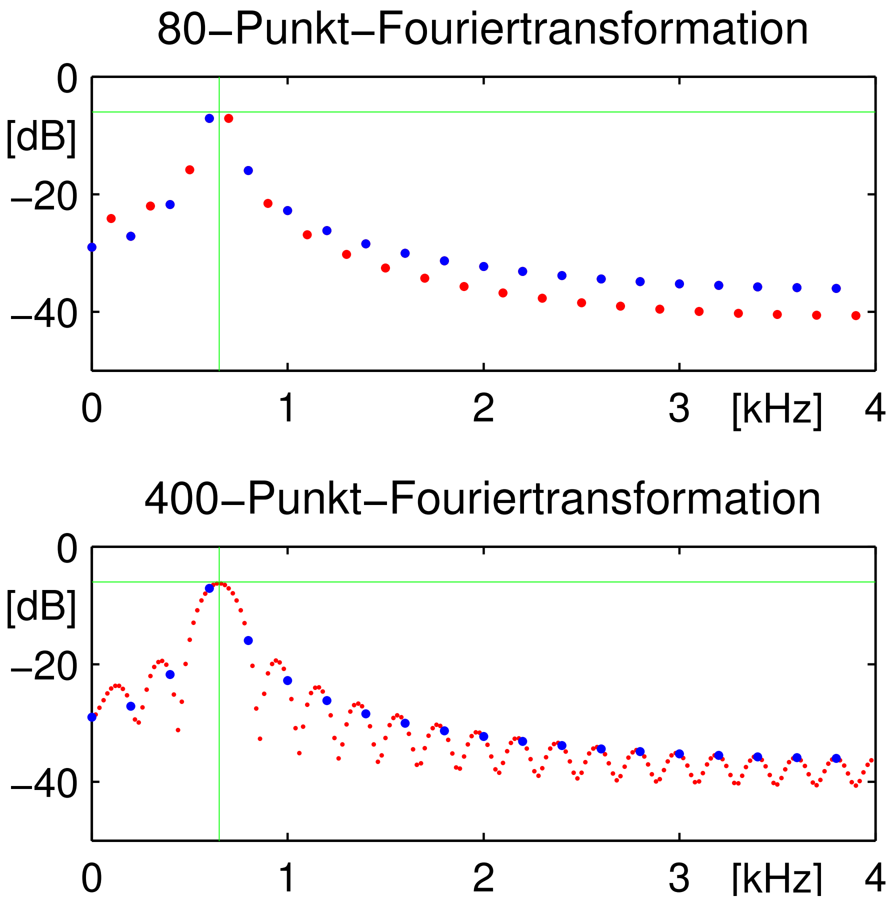
$$X'(2k) = \sum_{n=0}^{2N-1} x'(n) e^{-j(2\pi/(2N))2kn} = \sum_{n=0}^{N-1} x(n) e^{-j(2\pi/(2N))2kn}$$

$$= \sum_{n=0}^{N-1} x(n) e^{-j(2\pi/N)kn} = X(k)$$

Hochauflösende Fouriertransformation







80−Punkt−Fouriertransformation
0
[dB]
−20
−40
0
1
2
3
[kHz]
4
400−Punkt−Fouriertransformation
0
[dB]
−20
−40
0
1
2
3
[kHz]
4

Spektrum rauschartiger Signale

Bisher periodische Signale betrachtet!

Wie ist für rauschartige Signale vorzugehen?
(z.B. stimmlose Sprachsignal-Ausschnitte)

Leistungsdichtespektrum

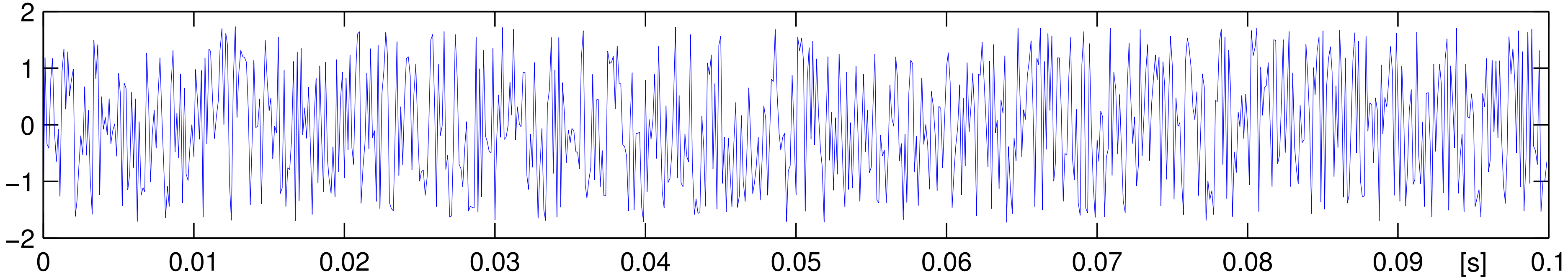
Schätzung des Leistungsdichtespektrum $S(\omega)$ für die stationäre Zufallssequenz $x(n)$ mittels DFT:

$$\tilde{S}(\omega_k) = \frac{1}{NU}|X(k)|^2 \qquad \text{mit} \qquad X(k) = \sum_{n=0}^{N-1} w(n)x(n)e^{-j(2\pi/N)kn}$$

$$\text{und} \qquad U = \frac{1}{N}\sum_{n=0}^{N-1} w^2(n)$$

Leistungsdichtespektrum eines Rauschsignals

Gegeben: Signal: weisses Rauschen mit 0 dB und $f_s = 8\,\text{kHz}$



Gesucht: Leistungsdichtespektrum

Methode:
- Kurzzeit-DFT >>>
- Kurzzeit-DFT mit Zero-Padding >>>
- Langzeit-DFT >>>
- Kurzzeit-DFT und Mittelung >>>

Spektralanalyse von Sprachsignalen

Stimmhafte Abschnitte:	Kurzzeit-DFT	Frequenzauflösung durch Länge des Analysefensters gegeben (ev. mit Zero-Padding falls Genauigkeit nötig)
Stimmlose Abschnitte:	Kurzzeit-DFT	Mittelung reduziert Varianz (besser cepstrale Glättung)

Autokorrelationsfunktion

für energiebegrenzte diskrete Signale $x(n)$:

$$r(k) = \sum_{m=-\infty}^{\infty} x(m)\, x(m+k)$$

für nicht energiebegrenzte (periodische und stationäre stochastische) diskrete Signale:

$$r(k) = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{m=-N}^{N} x(m)\, x(m+k)$$

AKF eines periodischen Signals ist auch periodisch!

Kurzzeit-Autokorrelationsfunktion

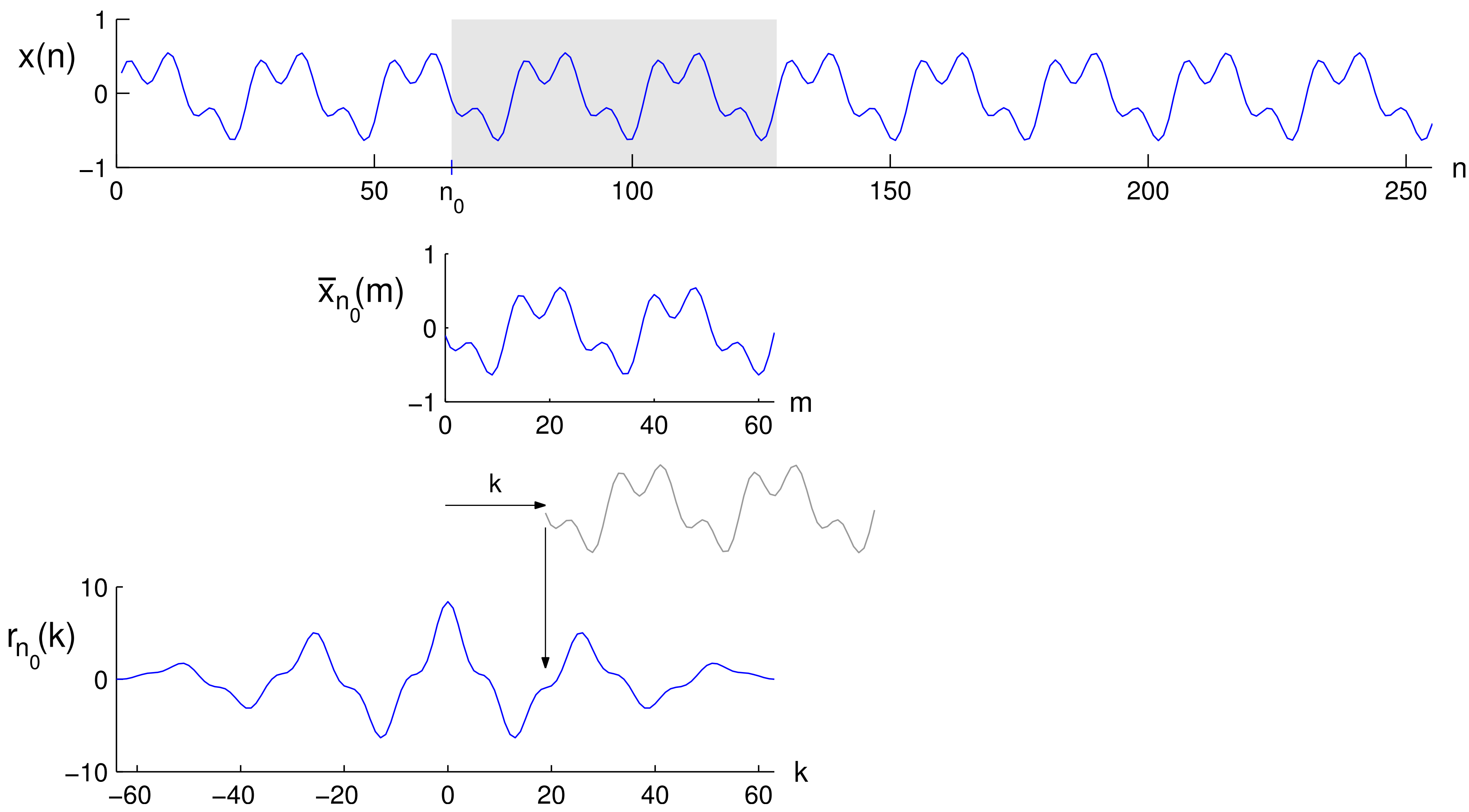
für die Sprachverarbeitung wichtiger (endliche Summe)

$$r_n(k) = \sum_{m=n}^{n+N-1} x(m)\, w(m-n)\, x(m+k)\, w(m-n+k) \qquad -N < k < N$$

oder äquivalent mit $\bar{x}_n(m) = x(n+m)\, w(m)$

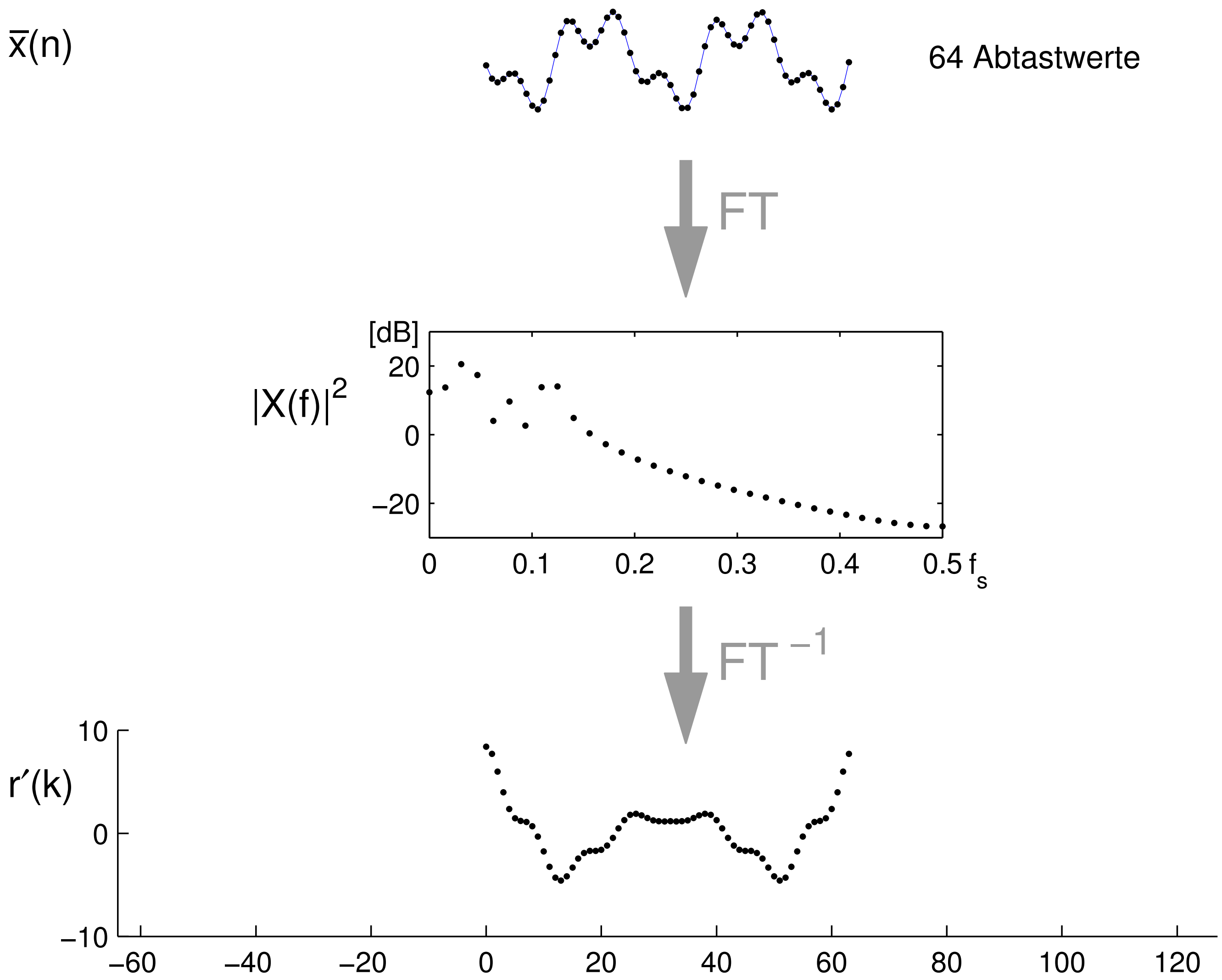
$$r_n(k) = \sum_{m=0}^{N-1-|k|} \bar{x}_n(m)\, \bar{x}_n(m+k) \qquad |k| < N$$

Achtung: Mit zunehmendem k werden weniger Terme aufsummiert, weil $\bar{x}_n(m)$ ausserhalb von $\{0, \ldots, N-1\}$ null ist!

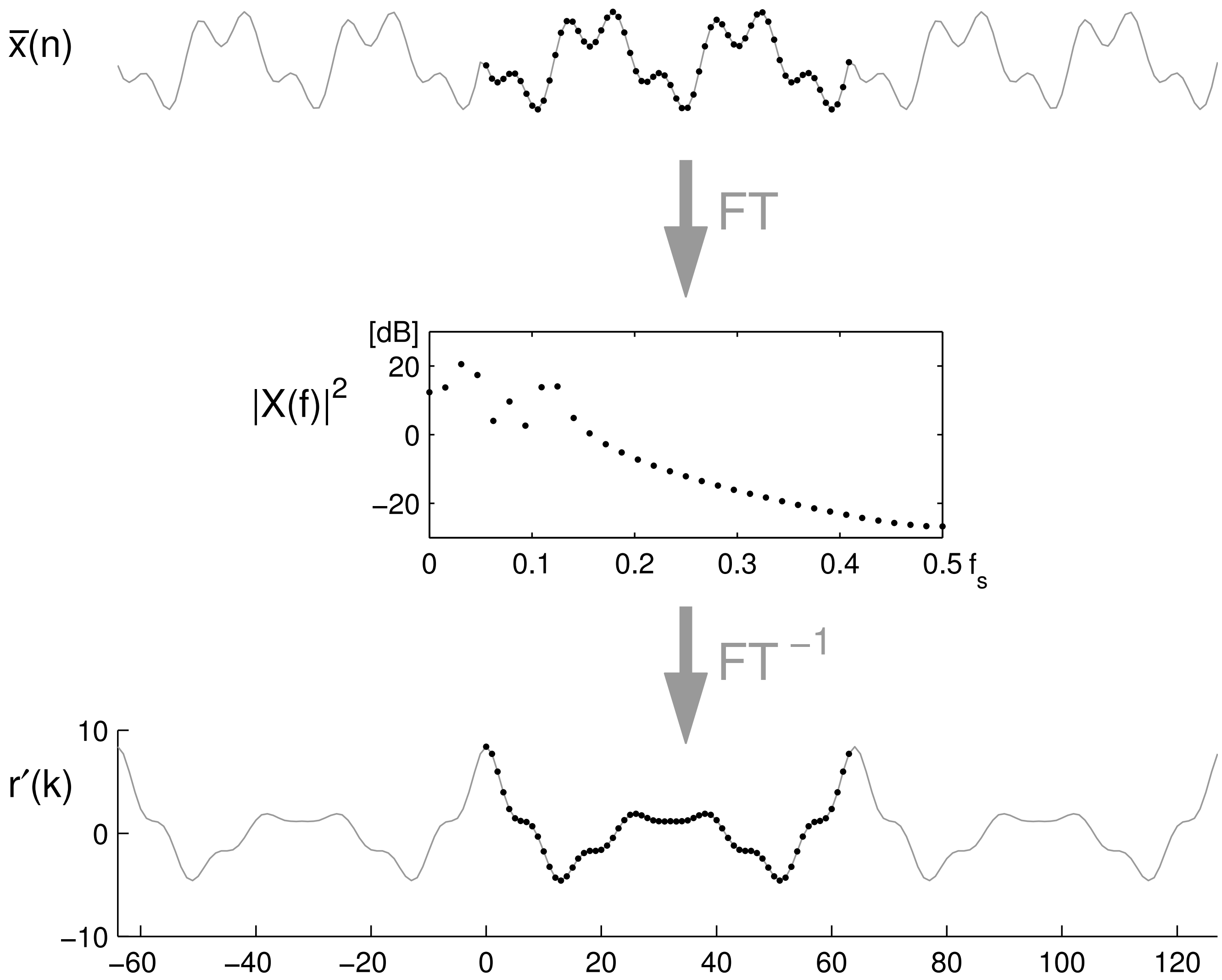

x(n)
1
0
–1
0
50
n_0
100
150
200
250
n
$\bar{x}_{n_0}(m)$
1
0
–1
0
20
40
60
m
k
$r_{n_0}(k)$
10
0
–10
–60
–40
–20
0
20
40
60
k

Effiziente Berechnung der AKF

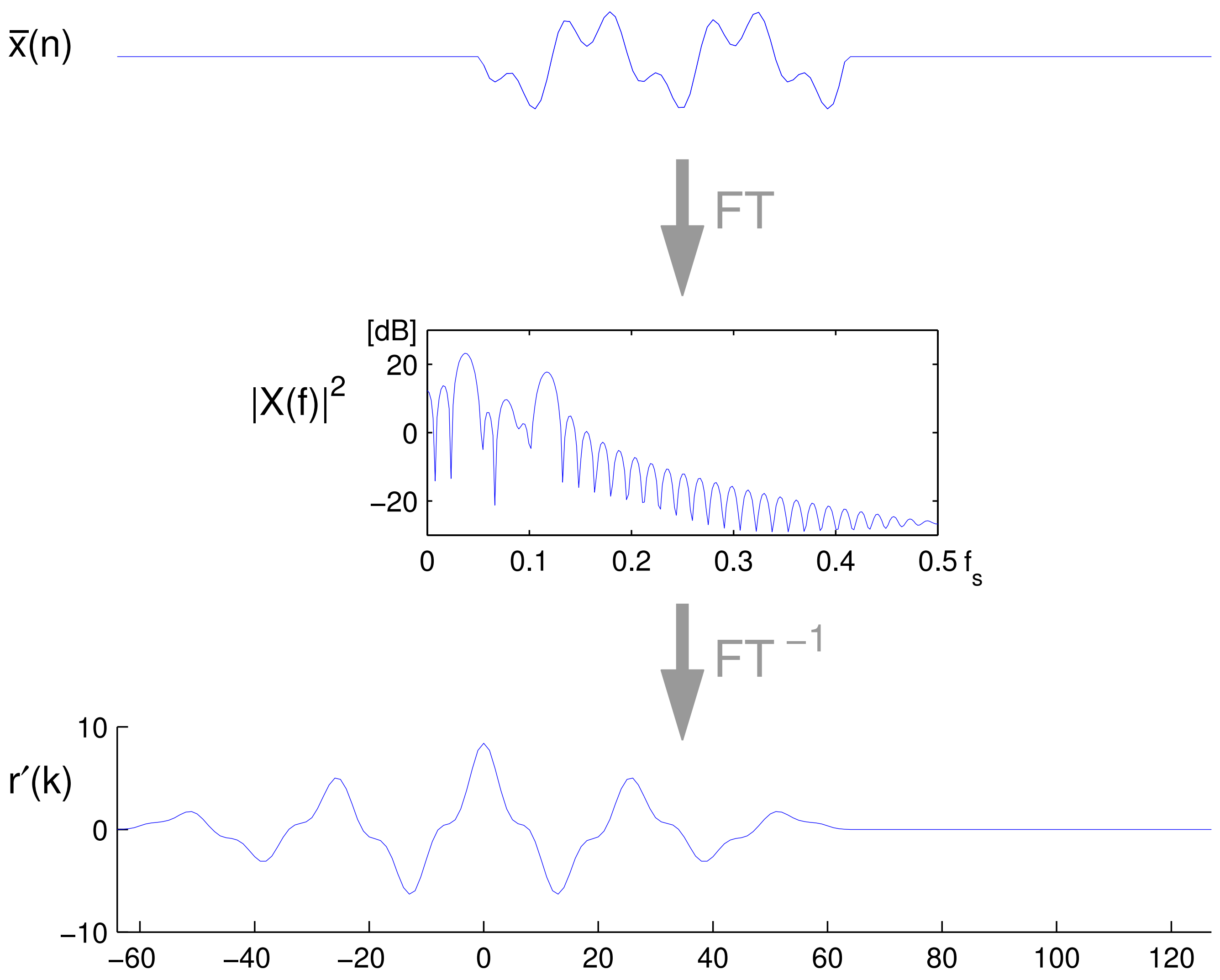
Feststellung:	AKF und Leistungsdichtespektrum bilden ein Fourierpaar
Folge:	AKF kann mittels DFT berechnet werden!
Frage:	Ist das Ergebnis gleich?



x̄(n)
64 Abtastwerte
FT
[dB]
20
0
−20
|X(f)|²
0
0.1
0.2
0.3
0.4
0.5 f_s
FT⁻¹
r′(k)
10
0
−10
−60
−40
−20
0
20
40
60
80
100
120

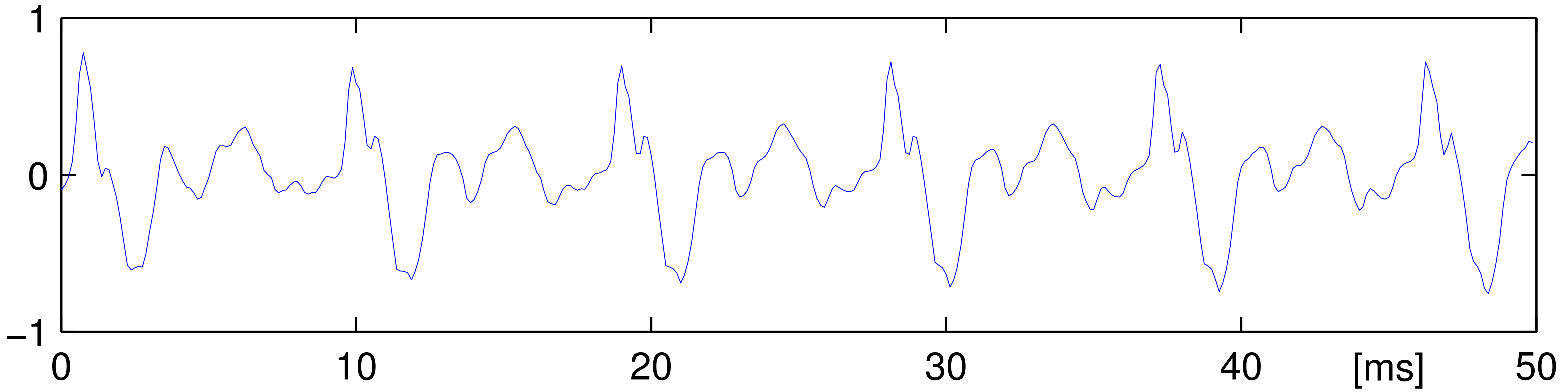


x̄(n)
FT
[dB]
20
0
−20
|X(f)|²
0
0.1
0.2
0.3
0.4
0.5 f
s
FT −1
10
0
−10
r′(k)
−60
−40
−20
0
20
40
60
80
100
120

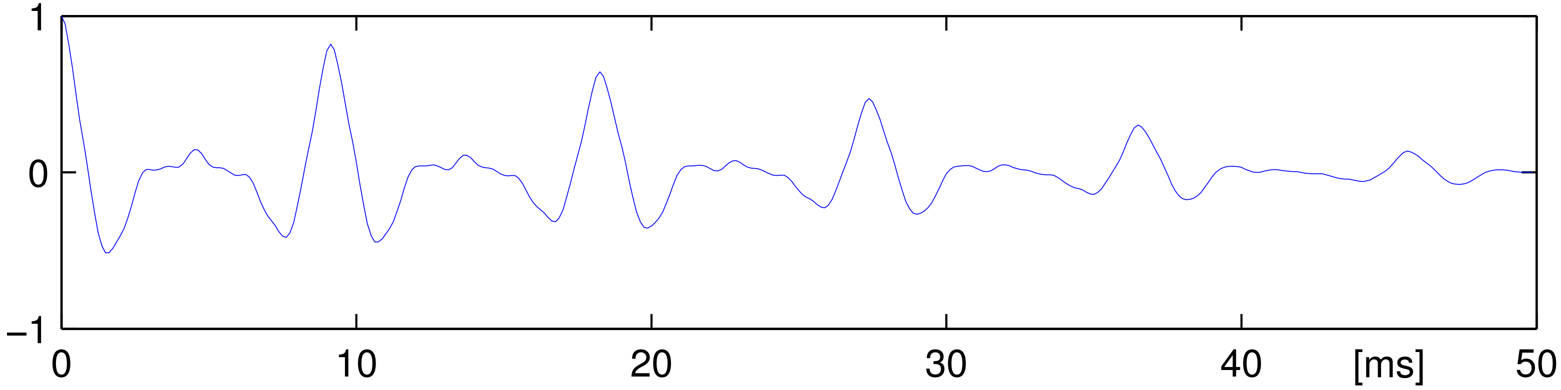


$\overline{x}(n)$
FT
[dB]
20
0
−20
$|X(f)|^2$
0
0.1
0.2
0.3
0.4
0.5 f_s
FT^{-1}
10
0
−10
$r'(k)$
−60
−40
−20
0
20
40
60
80
100
120

AKF eines stimmhaften Sprachsignalabschnittes

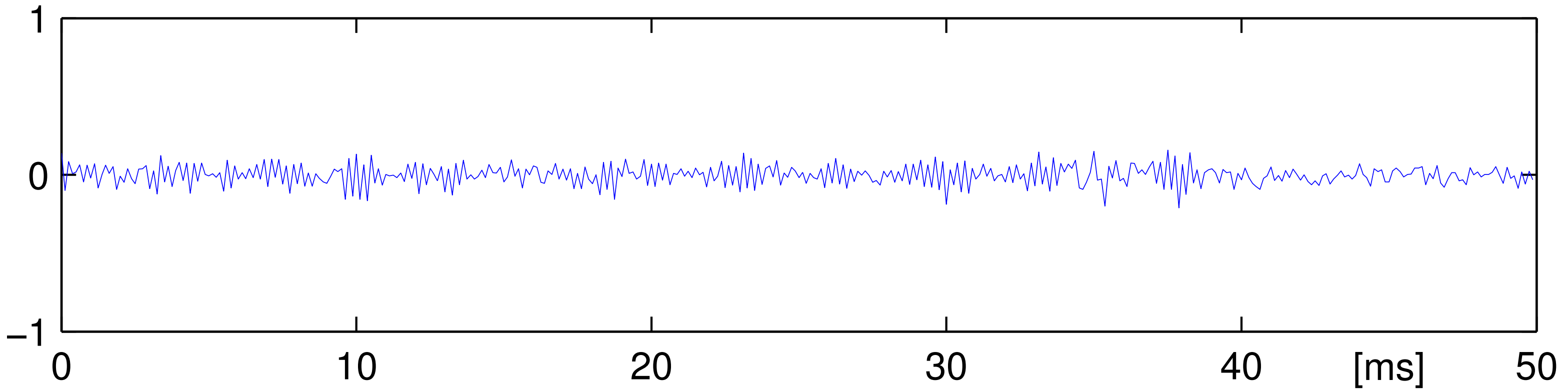


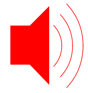
normierte AKF des stimmhaften Signalabschnittes



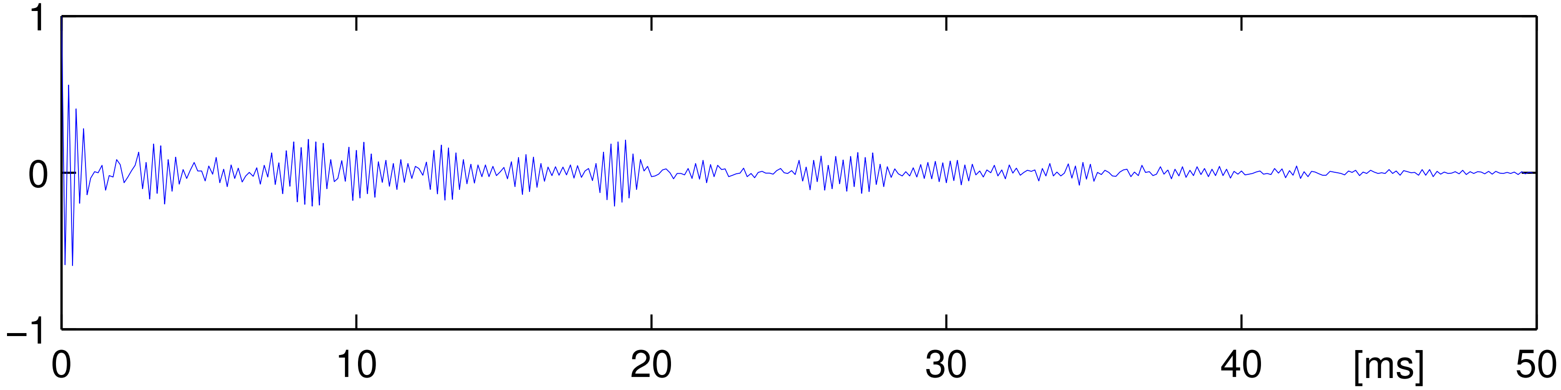
Position des 1. Maximums ⟶ Länge der Signalperiode

AKF eines stimmlosen Sprachsignalabschnittes





normierte AKF des stimmlosen Signalabschnittes



Kein ausgeprägtes Maximum $\longrightarrow$ Signal nicht periodisch

Zusammenfassung

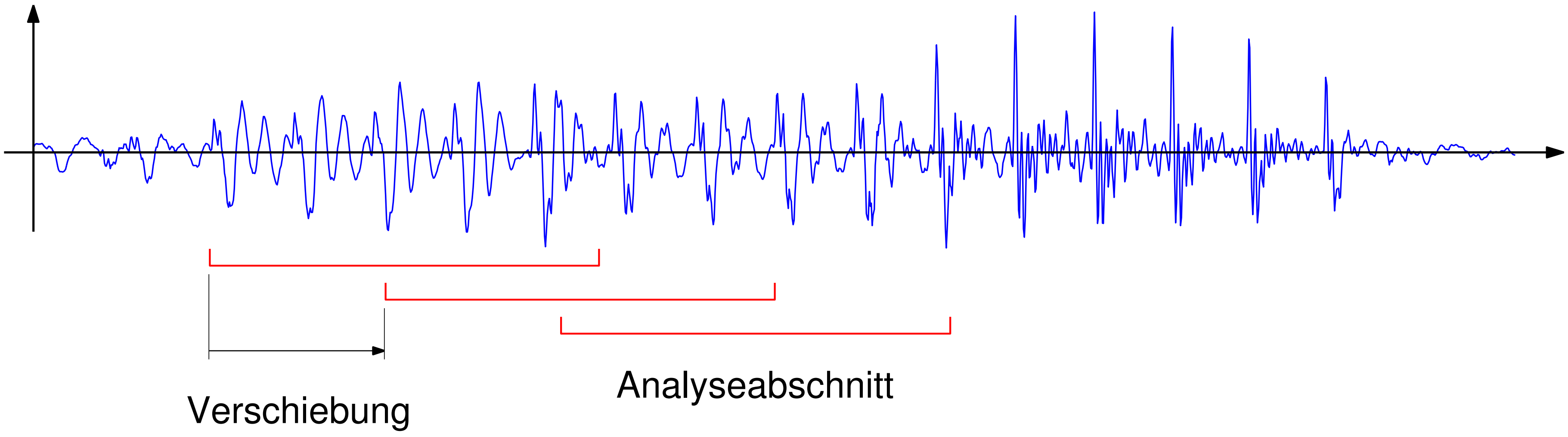
- Kurzzeitanalyse für quasi-stationäre Signale
- Fouriertransformation:
  - Fensterfuntion (Schmieren und Lecken)
  - hochauflösende FT (*zero padding*)
  - Spektrum nicht periodischer Signale
- Autokorrelationsfunktion:
  - Berechnung direkt oder via FT
  - Ermitteln der Signalperiode

Thema der nächsten Lektion:

Lineare Prädiktion

Zur Übersicht der Vorlesung *Sprachverarbeitung I* >>>

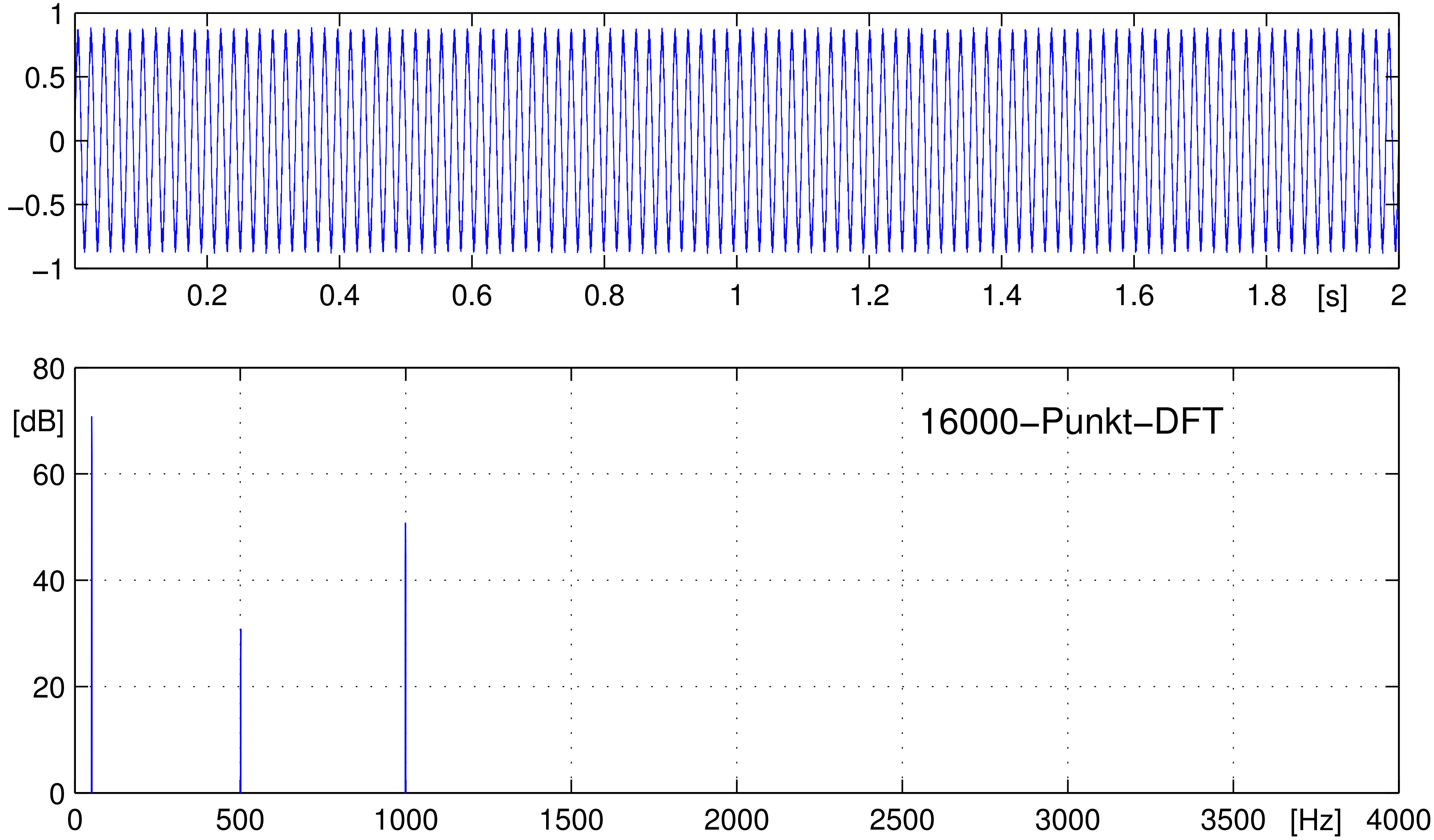
Kurzzeitanalyse quasi-stationäre Signale



Verschiebung des Analysefensters: T_a

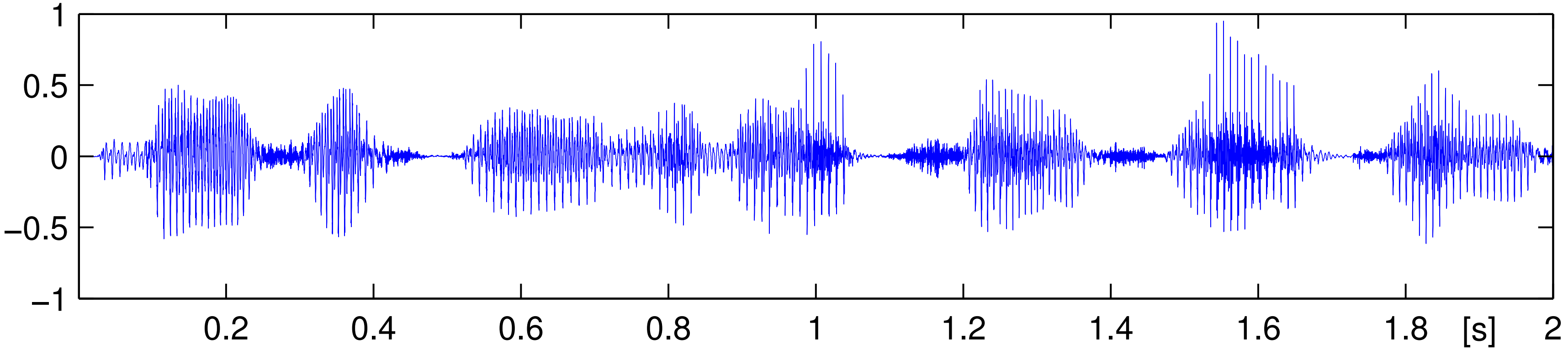
Analysefrequenz oder -rate: $F_a = 1/T_a$ <<<

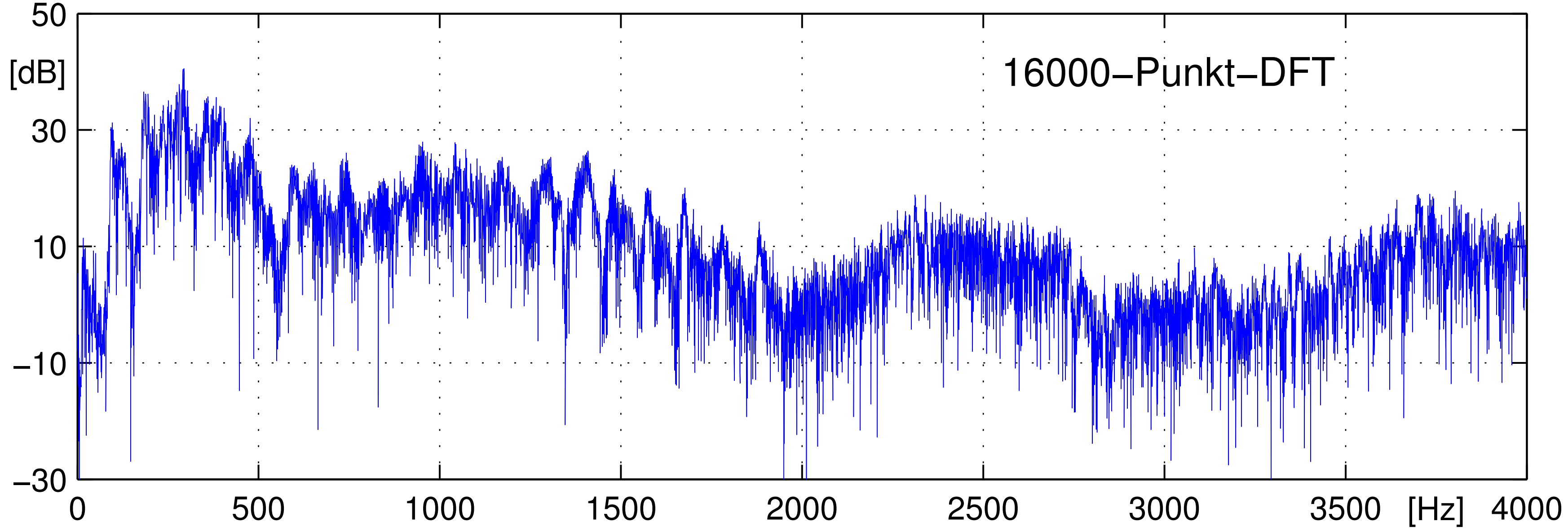
Langzeit-Spektrum stationäres Signal



<<<

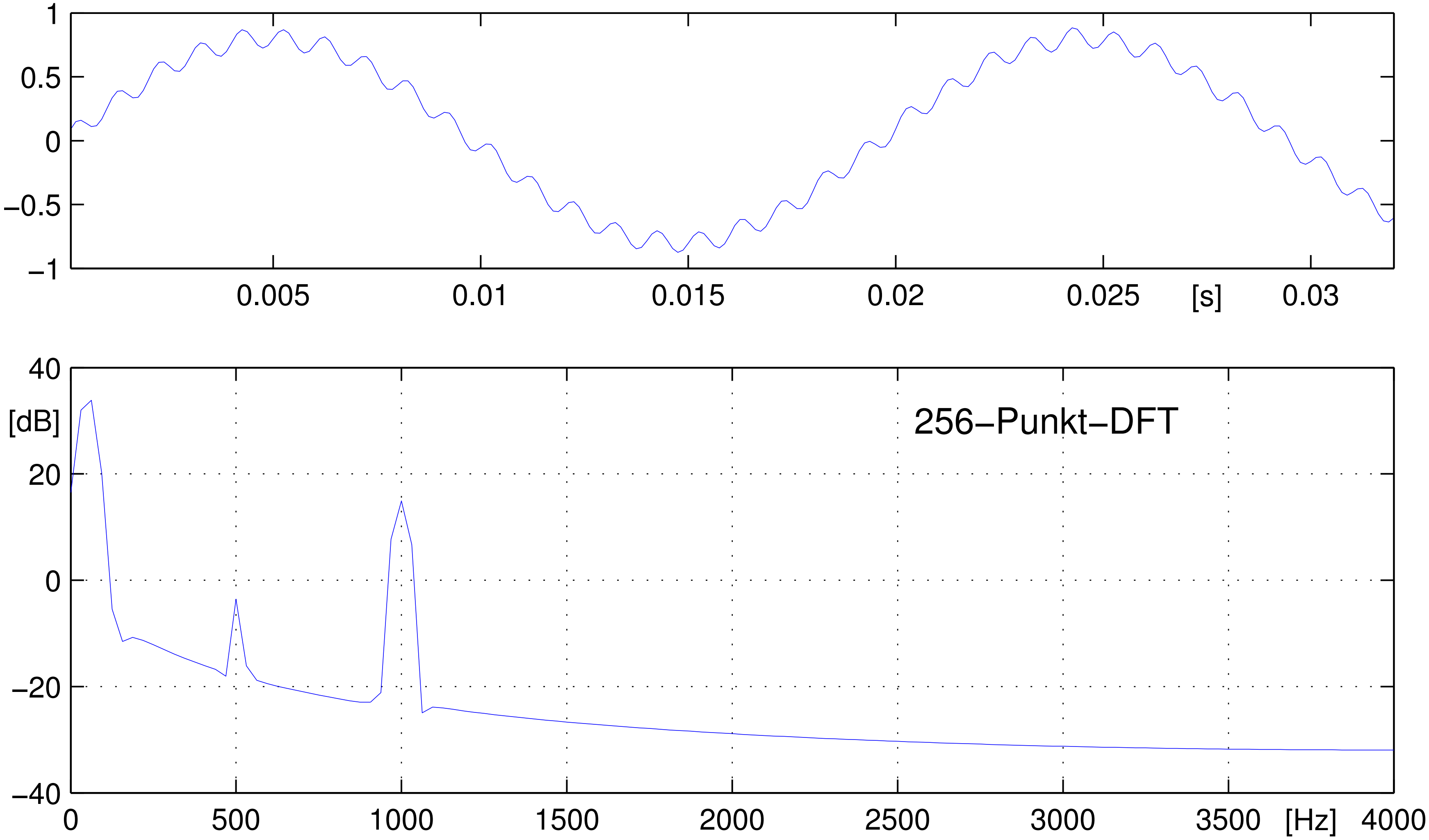
Langzeit-Spektrum Sprachsignal





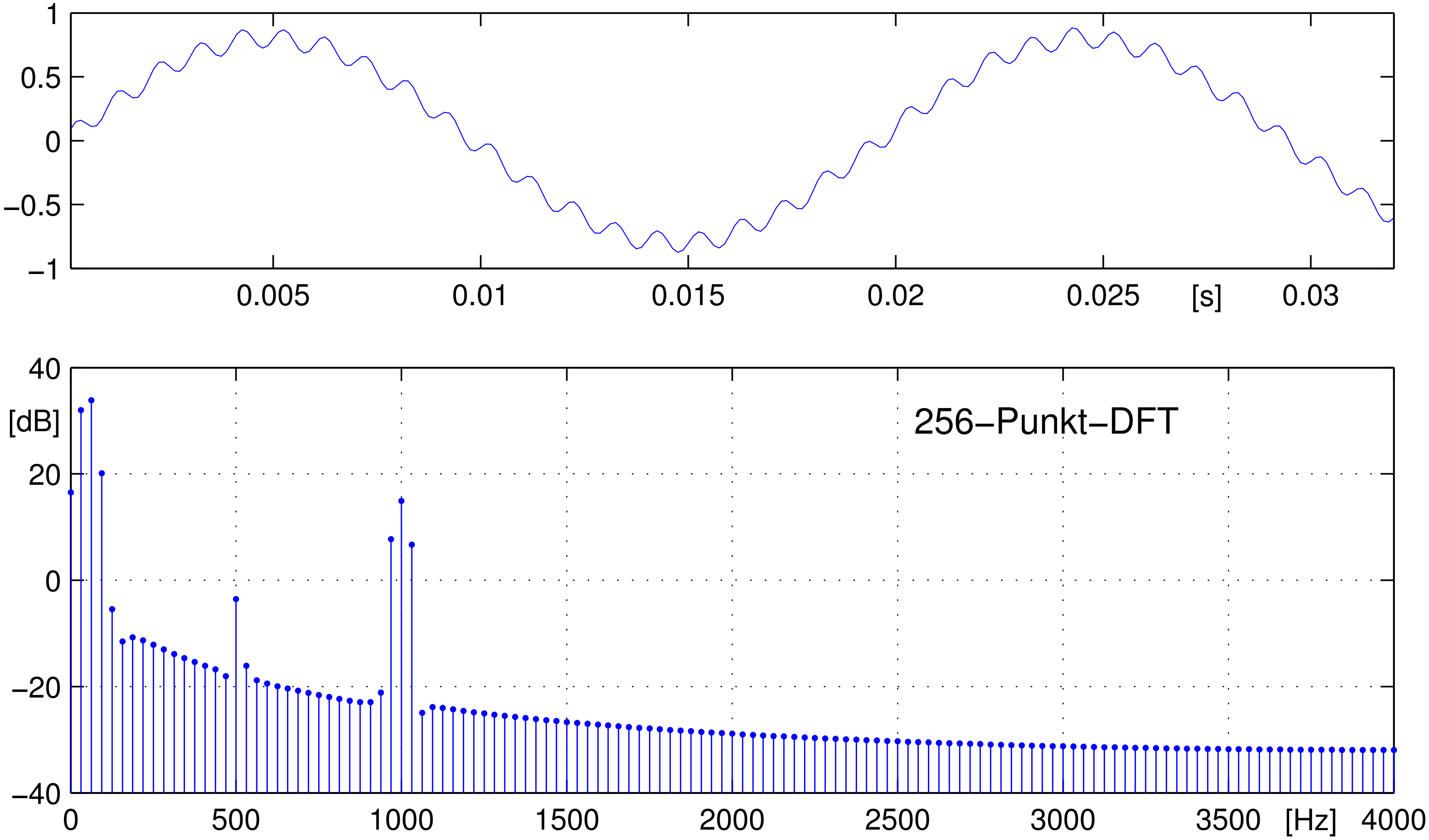
<<<

Kurzzeit-Spektrum stationäres Signal



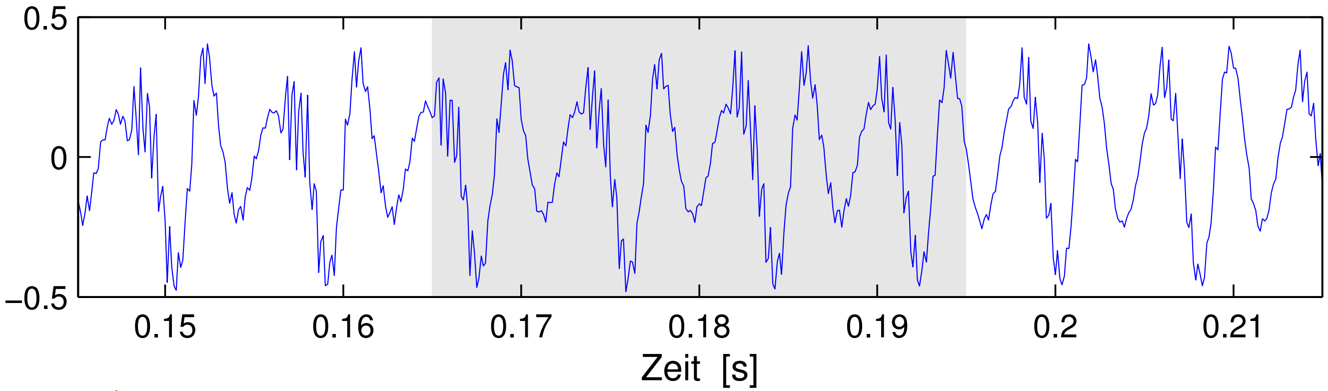
<<<

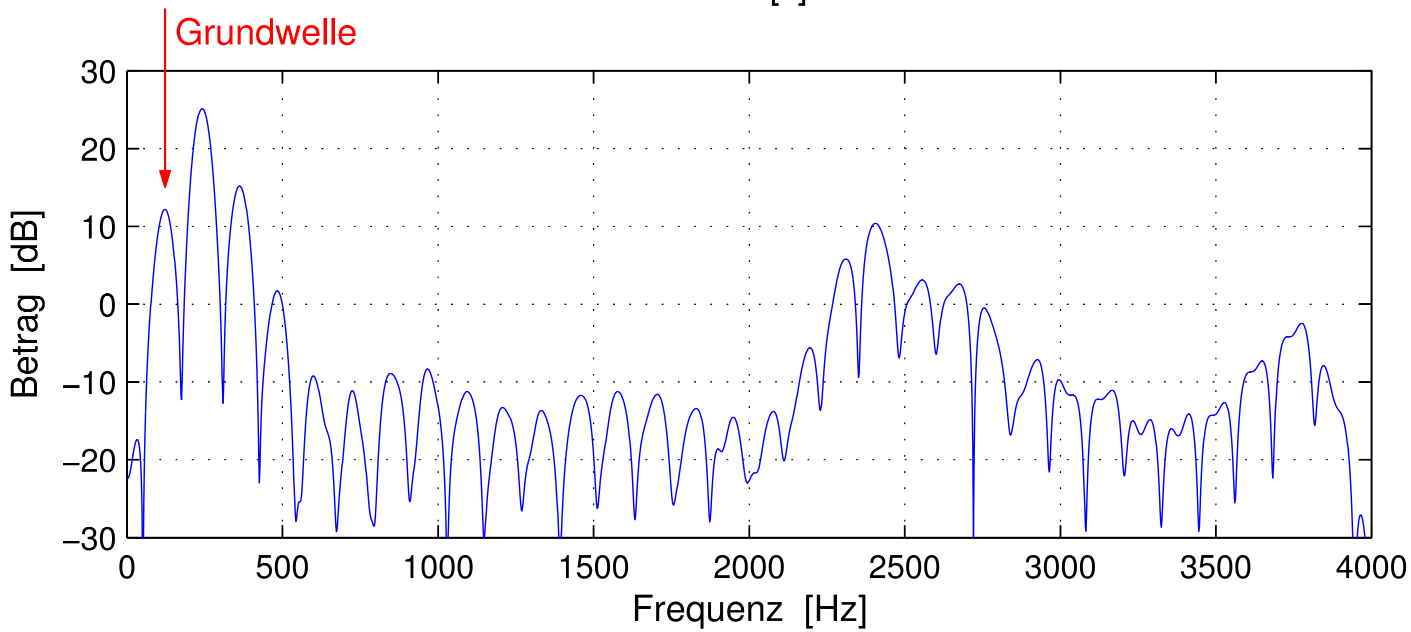
Kurzzeit-Spektrum stationäres Signal



<<<

Spektrum des Sprachsignals (stimmhafter Ausschnitt)

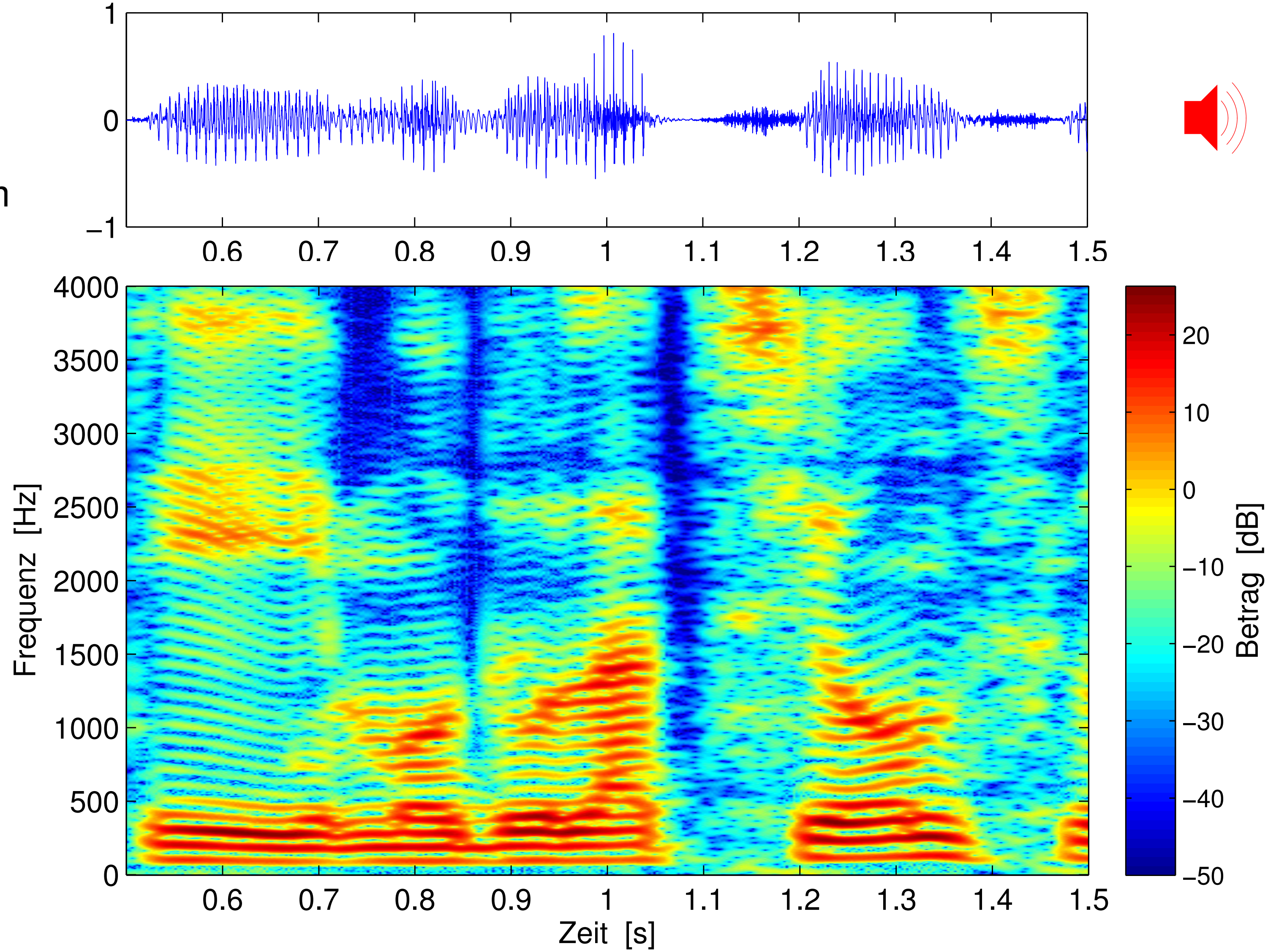




<<<

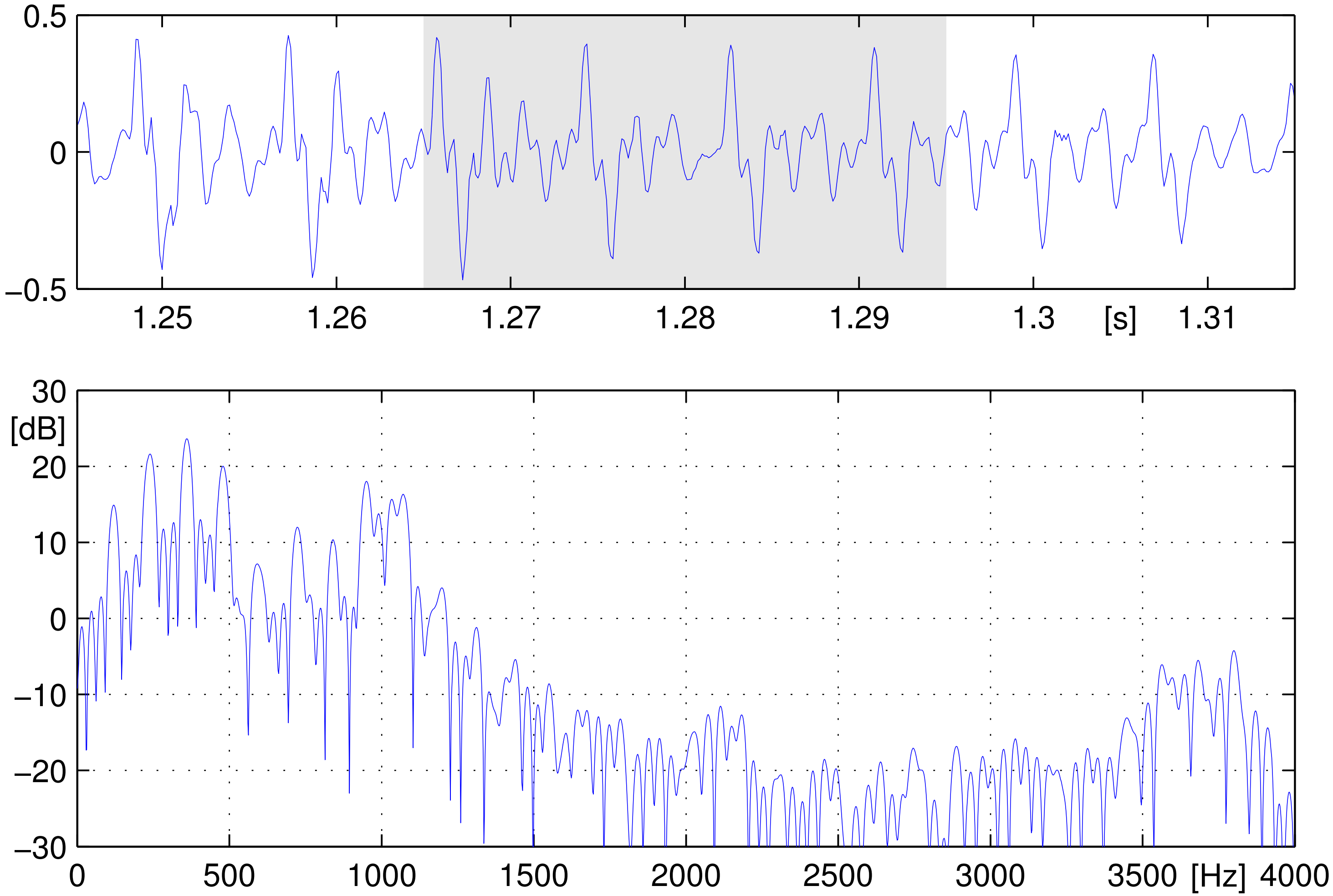
Spektrogramm

Schmalbandspektrogramm



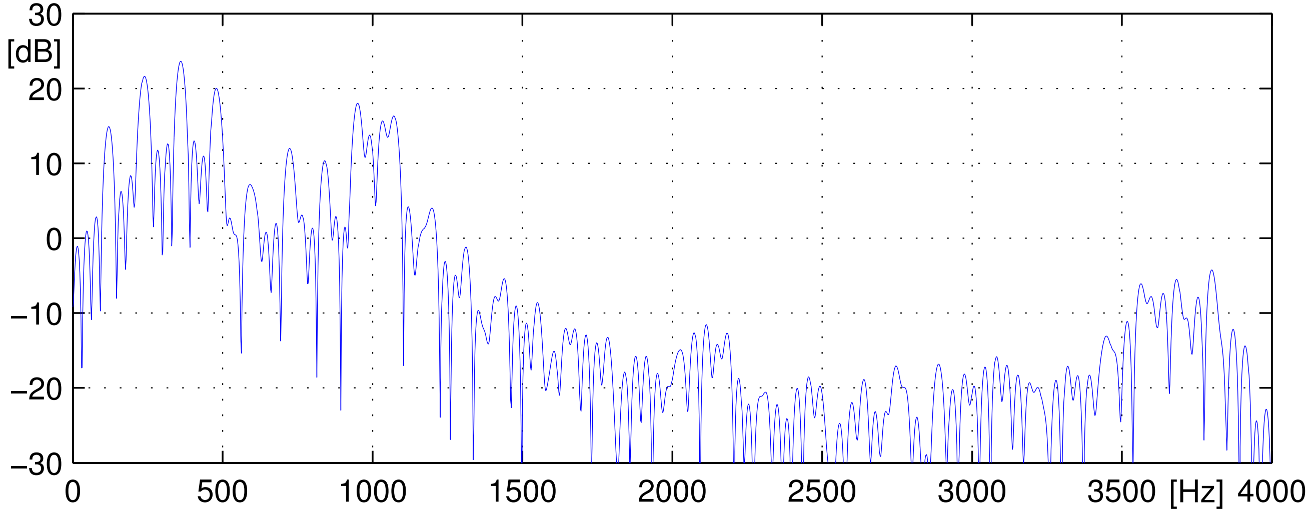
<<<

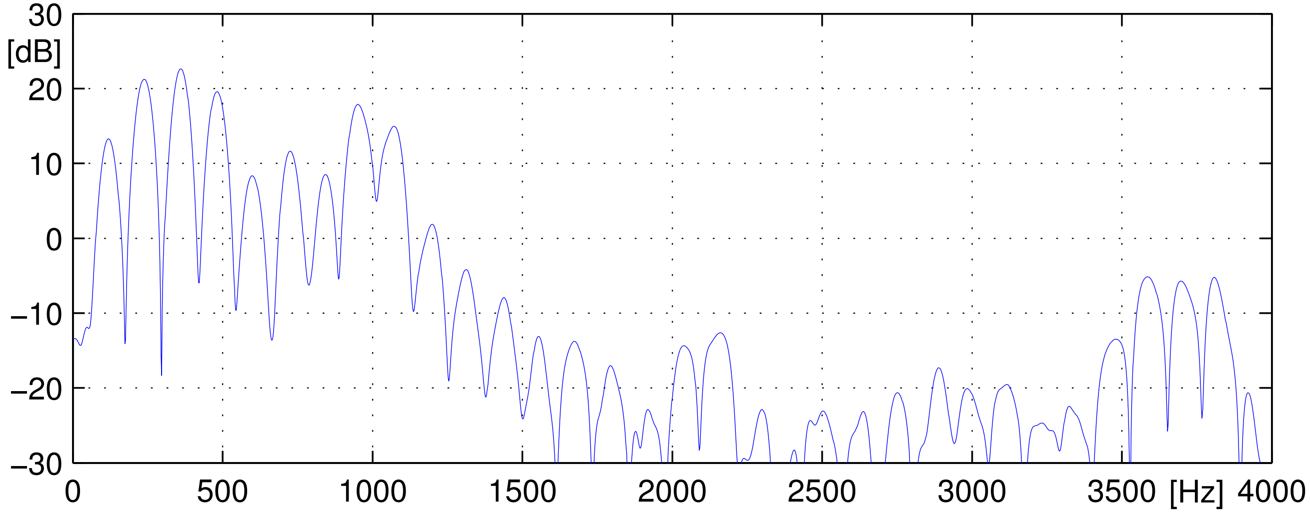
DFT mit Rechteckfensterfunktion



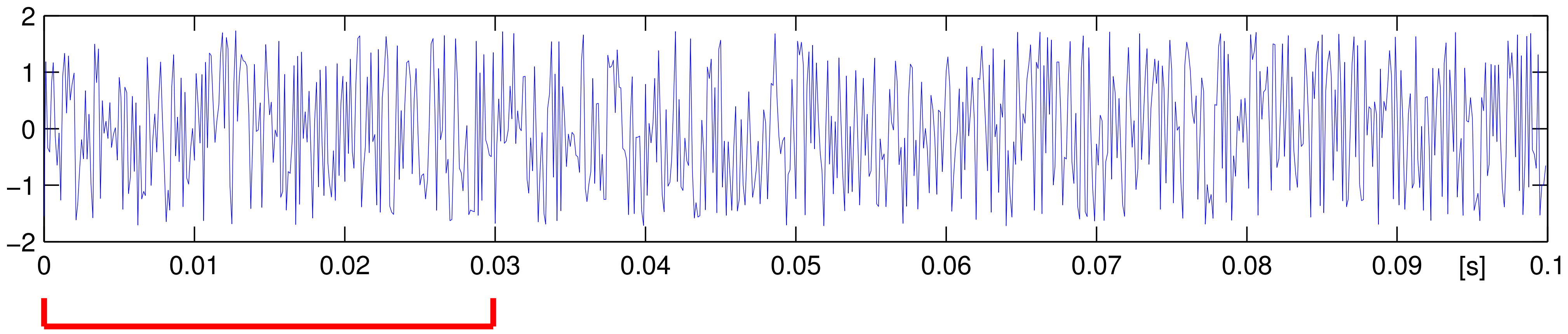
<<<

DFT mit Rechteck- bzw. Hamming-Fensterfunktion



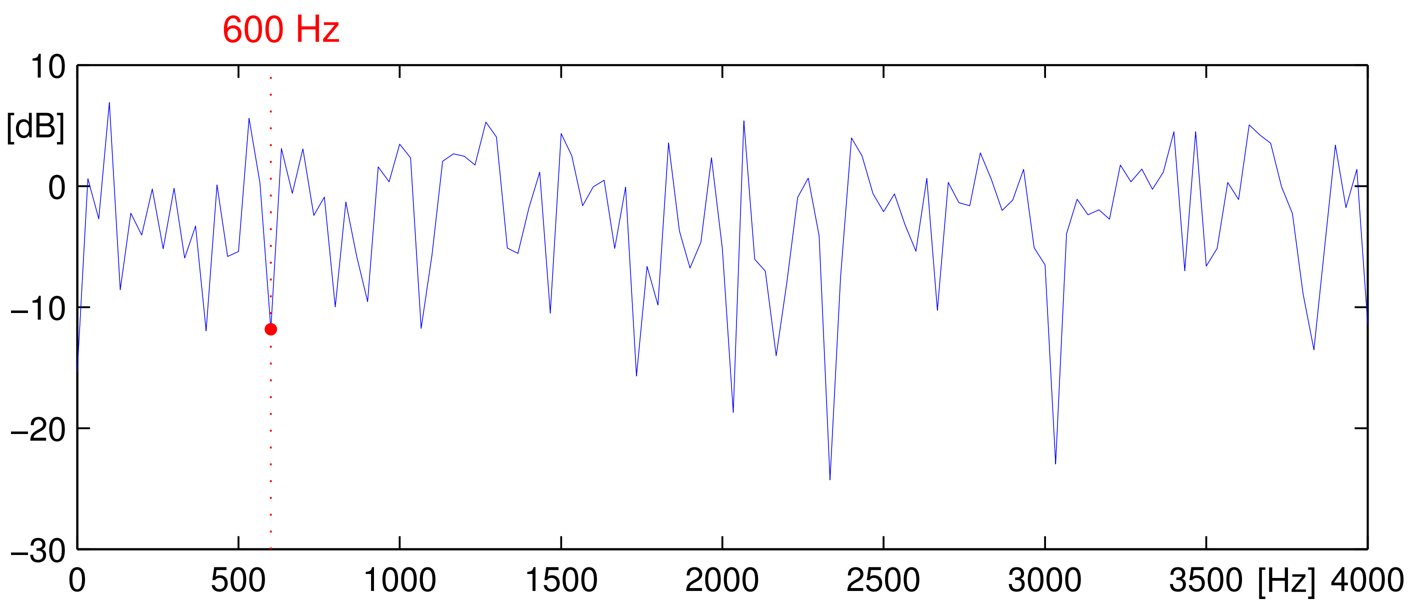


<<<

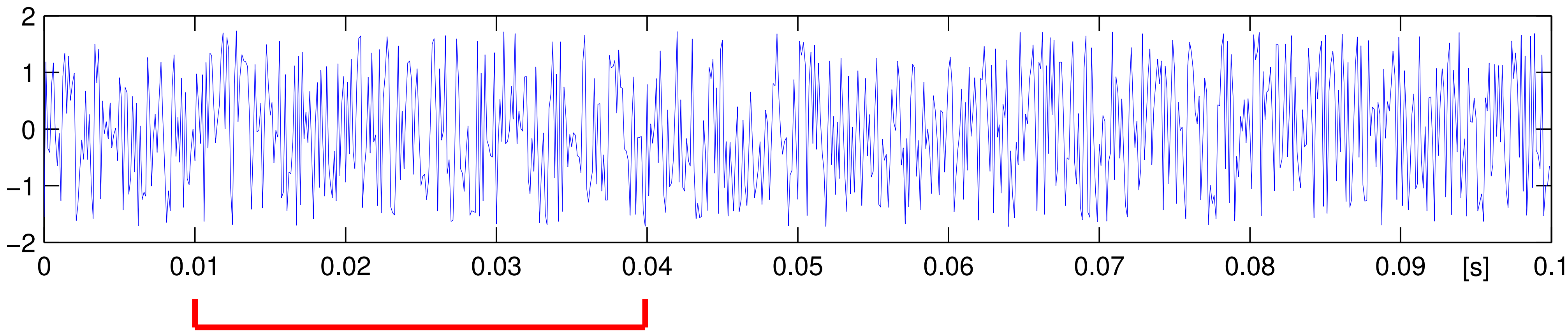


240-Punkt-DFT

(30 ms)

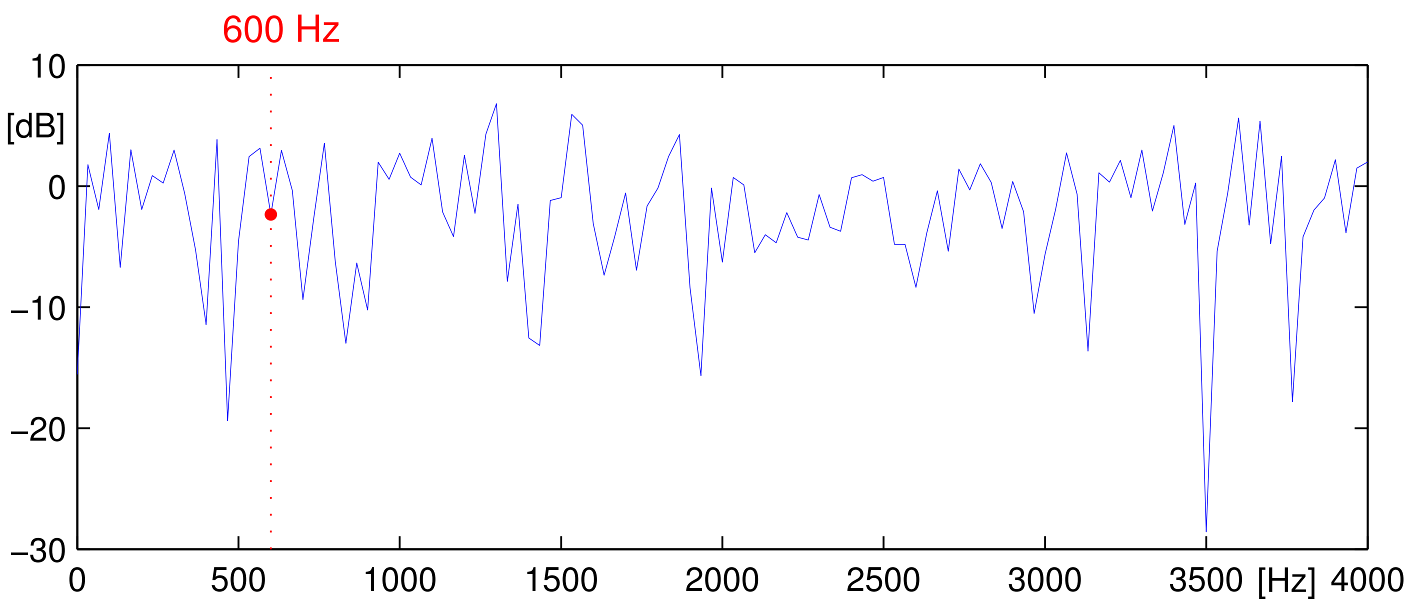


<<<

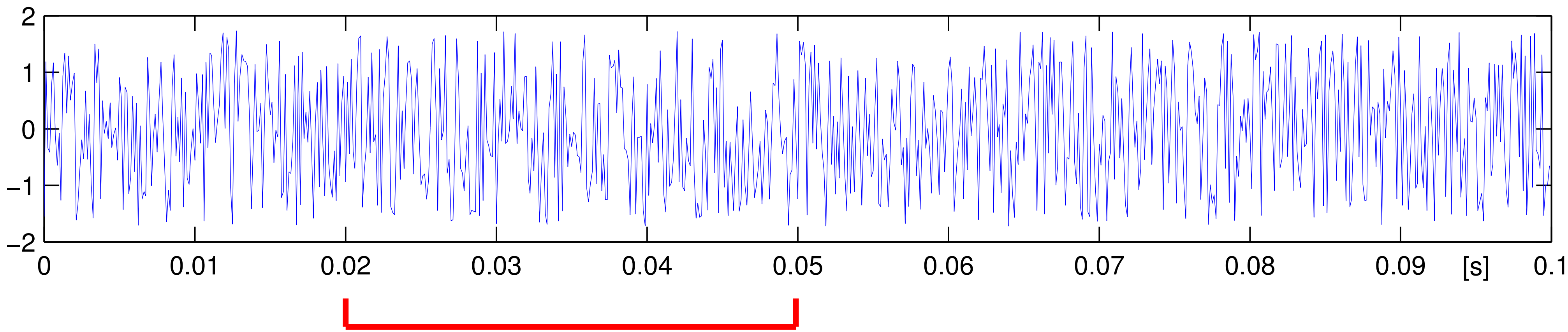


240-Punkt-DFT

(30 ms)

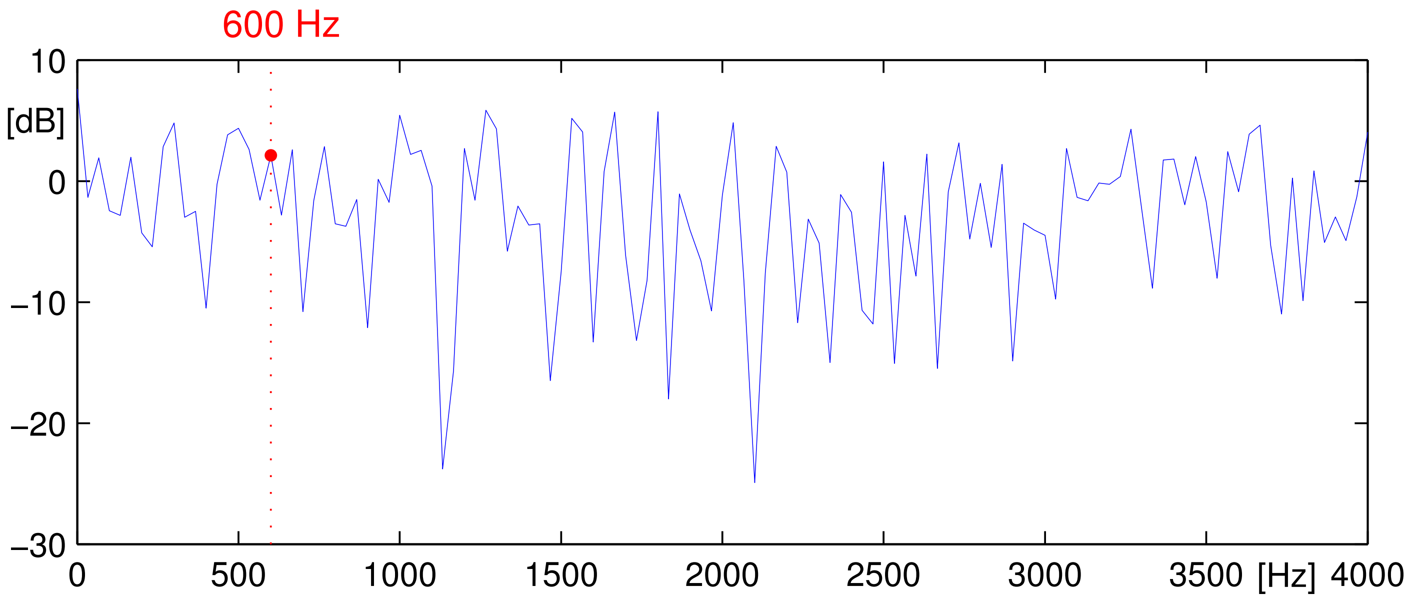


<<<

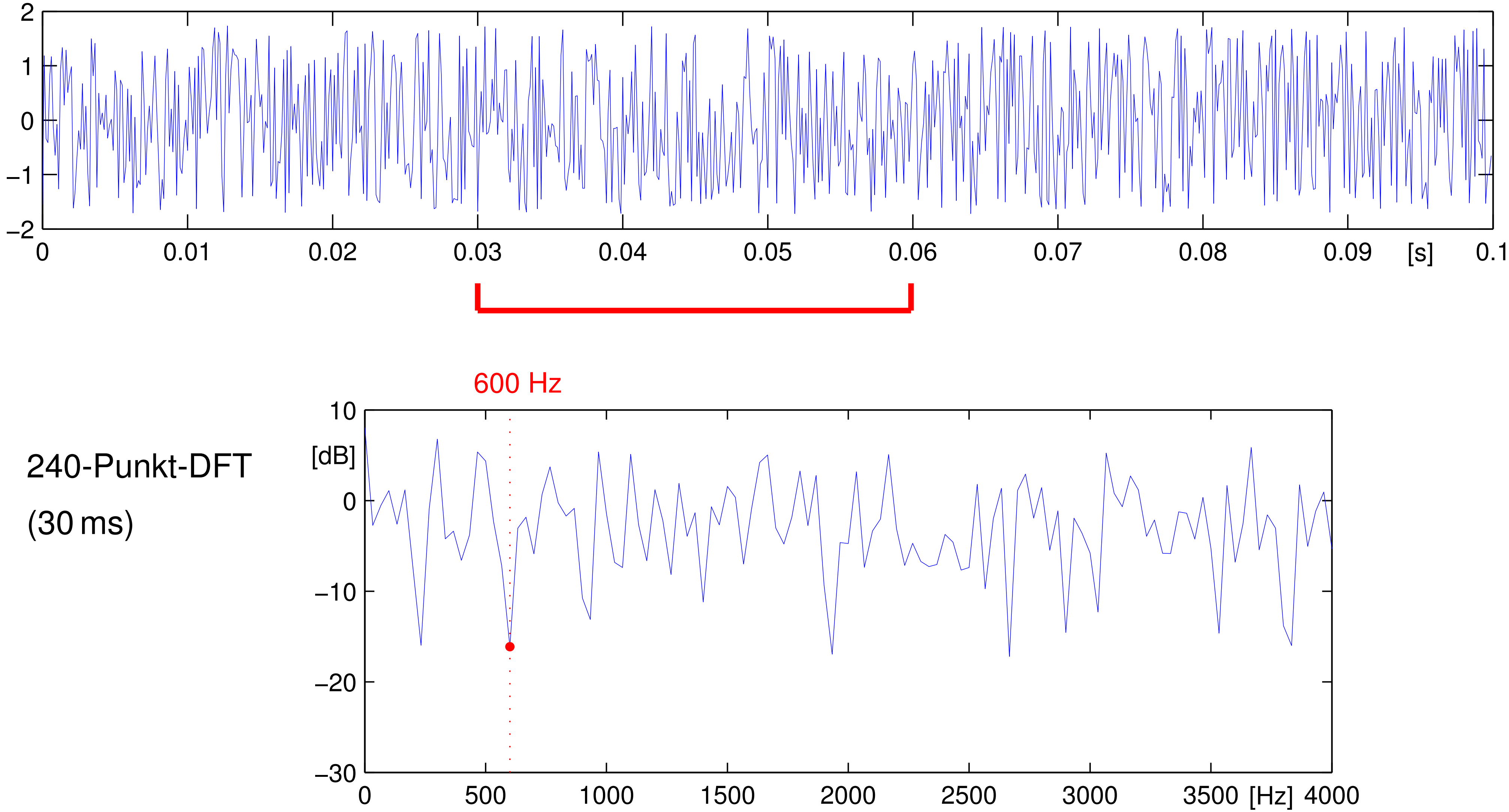


240-Punkt-DFT

(30 ms)



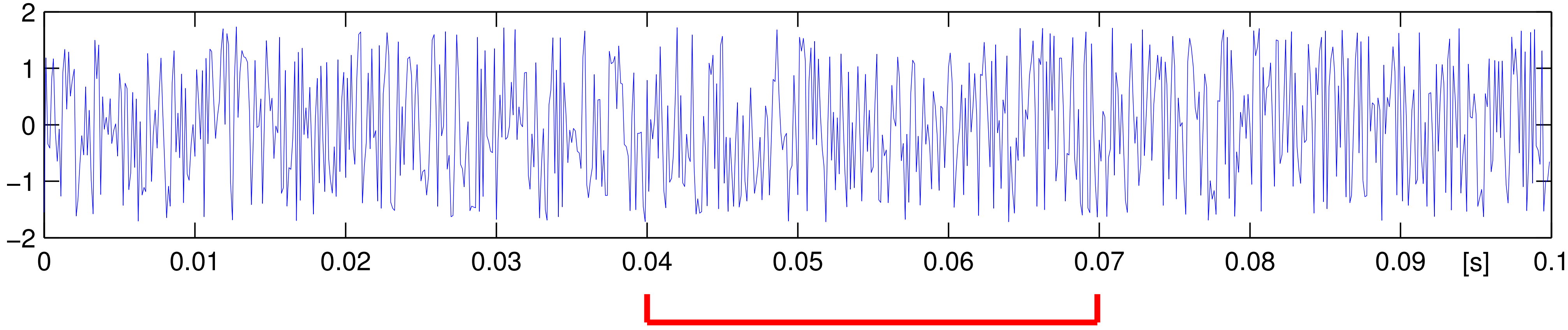
<<<



240-Punkt-DFT

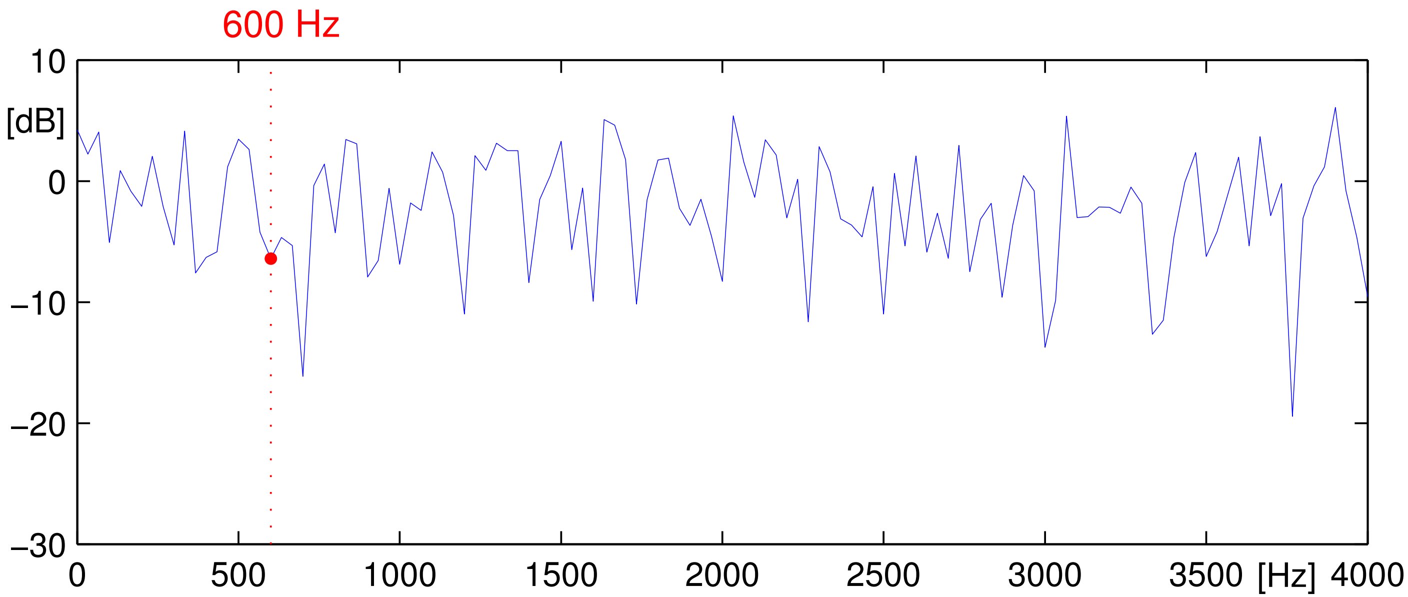
(30 ms)

<<<

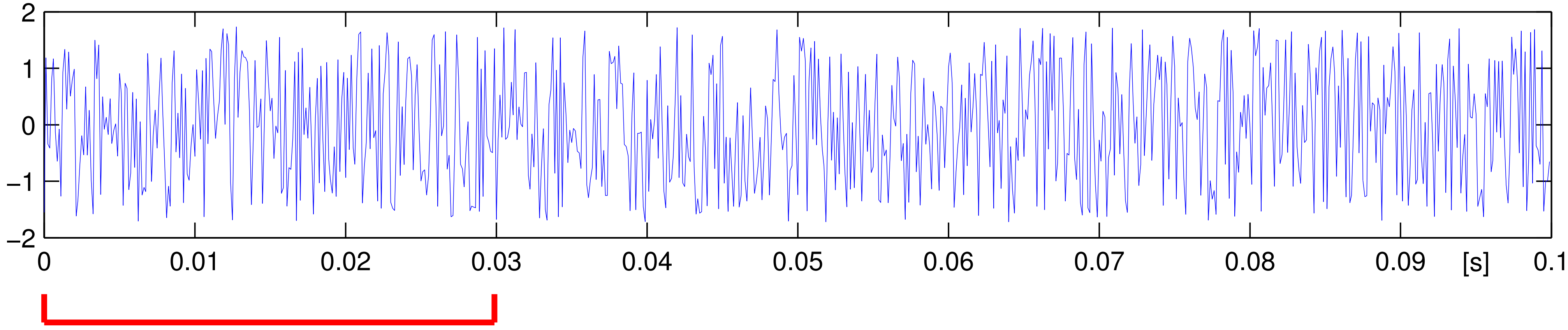


240-Punkt-DFT

(30 ms)



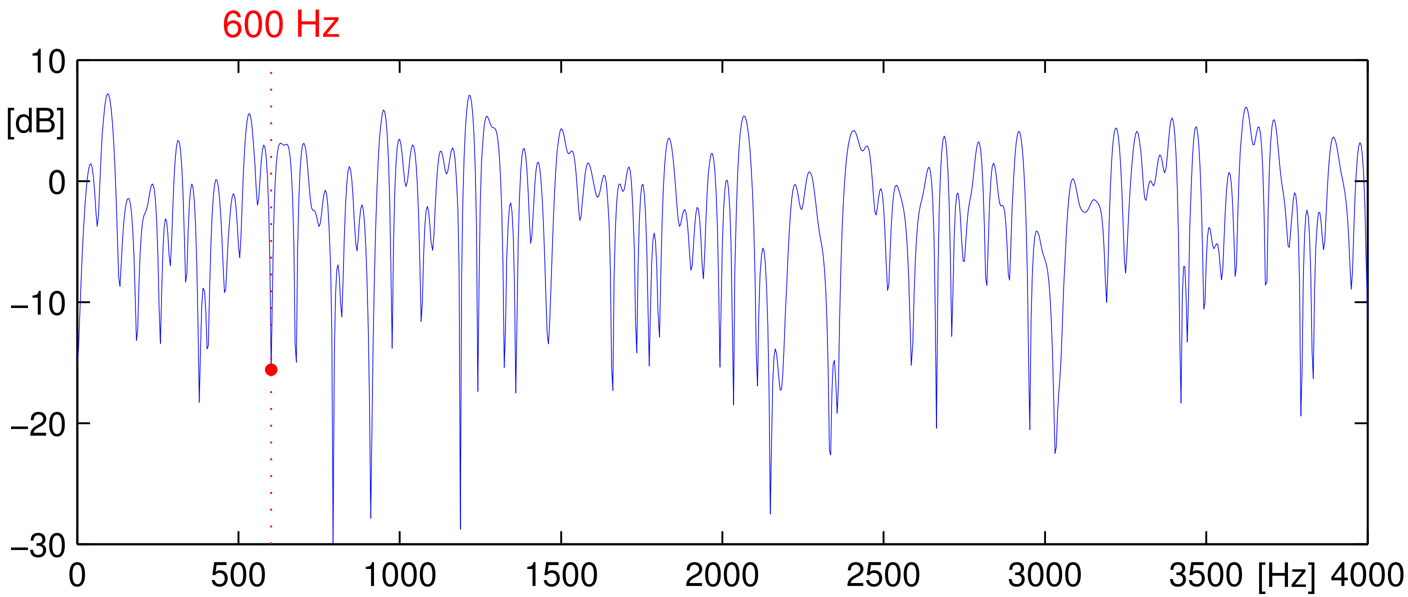
<<<



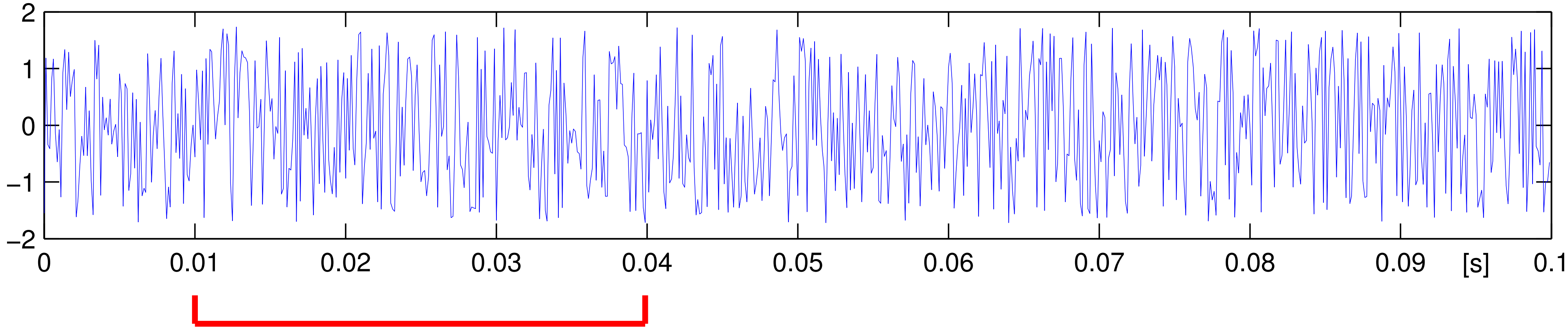
30 ms-Fenster

(240 Abtastwerte)

2048-Punkt-DFT



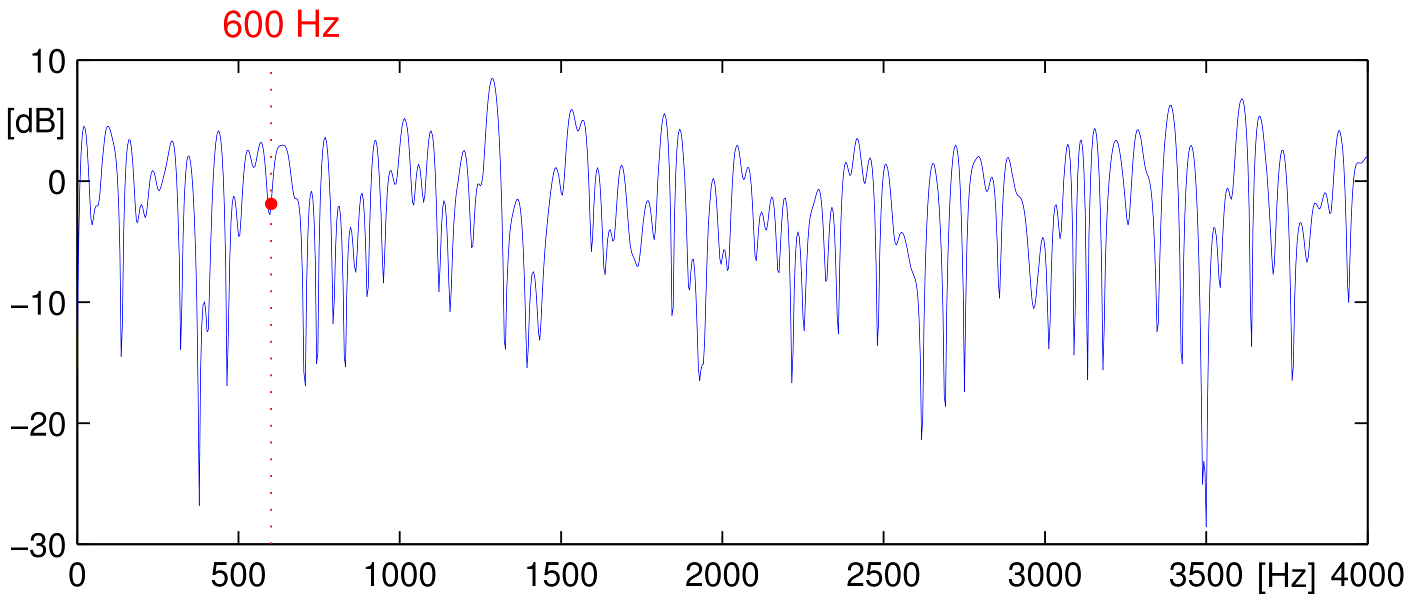
<<<



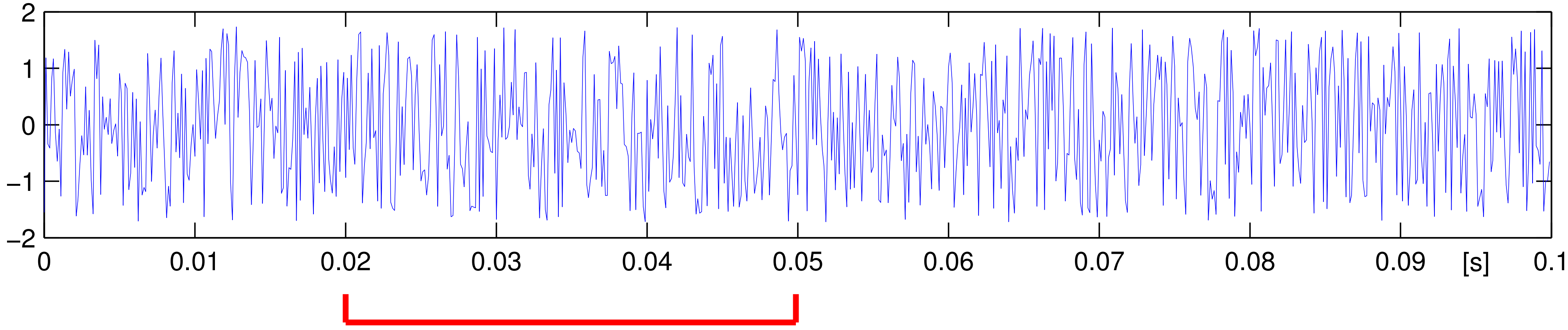
30 ms-Fenster

(240 Abtastwerte)

2048-Punkt-DFT



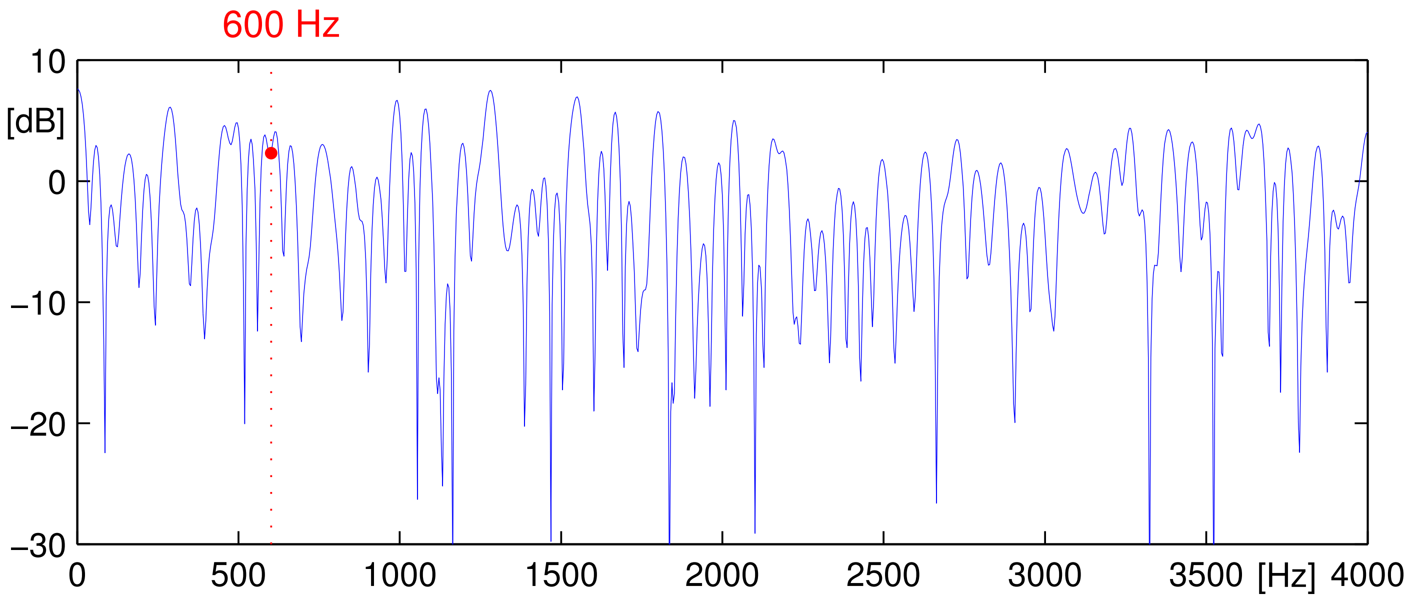
<<<



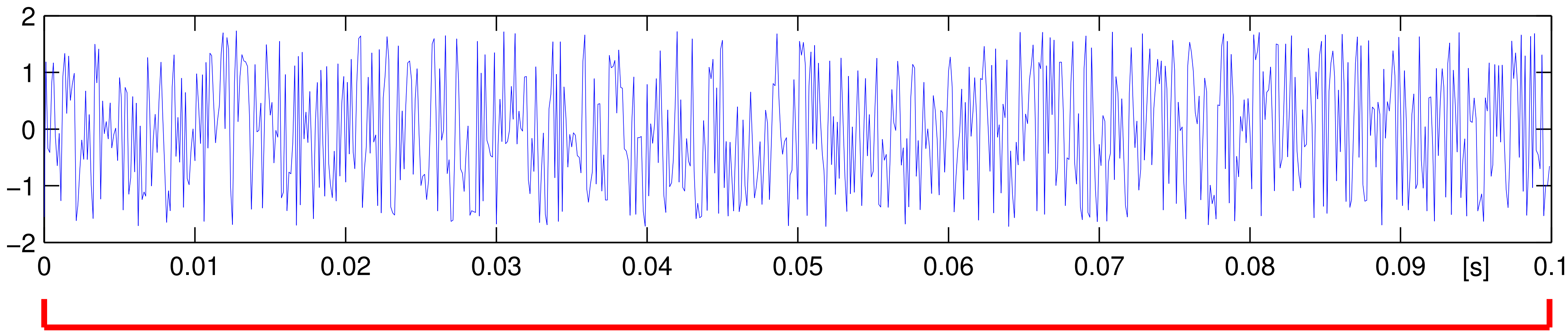
30 ms-Fenster

(240 Abtastwerte)

2048-Punkt-DFT

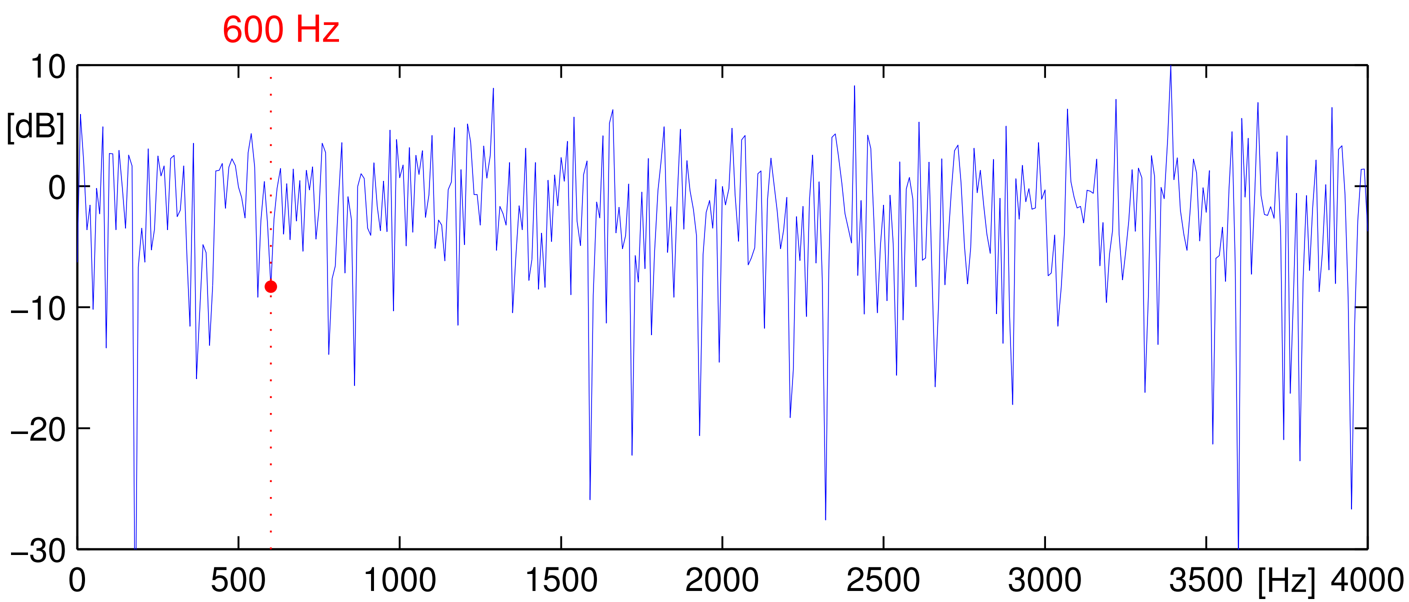


<<<



800-Punkt-DFT

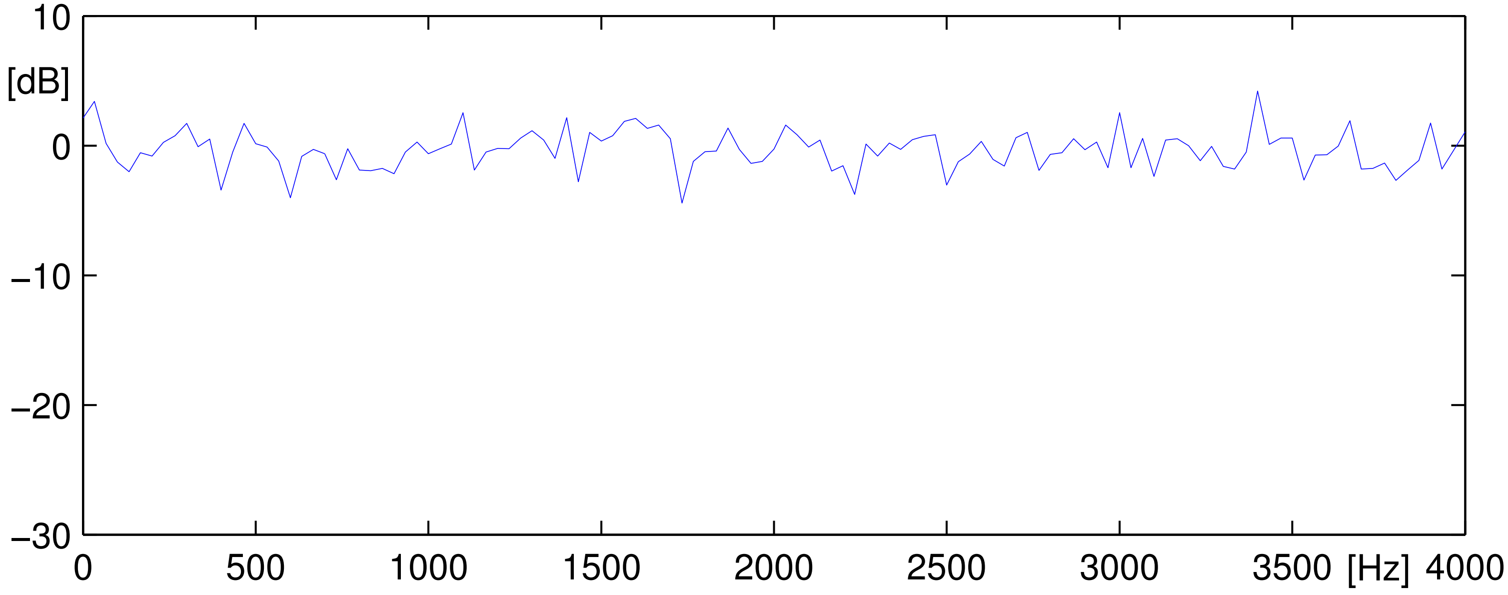
(100 ms)



<<<

Leistungsdichtespektrum eines Rauschsignals:

Mittelung über 10 Analysefenster (240-Punkt-DFT)



<<<